Chapitre VI

Méthodes Itératives – Equations Non Linéaires

En pratique, on est souvent confronté à la résolution d'un système d'équations non linéaires. C'est-à-dire pour une fonction $f : I\!R^n \to I\!R^n$ donnée, on cherche un point $x \in I\!R^n$ tel que

$$f(x) = 0. \tag{0.1}$$

En général, il n'y a pas d'algorithme fini pour trouver une solution. On est donc obligé d'utiliser des méthodes itératives.

Sans hypothèses supplémentaires, on ne sait rien sur l'existence d'une solution de (0.1). Par exemple, pour $f(x) = e^x$ il n'y a pas de solution, pour $f(x) = \sin x$ il y en a une infinité. Mais, le *théorème d'inversion locale* nous fournit un résultat sur l'unicité locale: si $f(a) = 0$ et si la matrice jacobienne $f'(a)$ est inversible, il existe un voisinage U de a et un voisinage V de 0 tels que $f : U \to V$ soit bijective. Ceci implique que a est la seule solution de (0.1) dans le voisinage U de a.

Bibliographie sur ce chapitre

P. Deuflhard (2004): *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*. Springer Ser. Comp. Math. 35, Springer, Berlin.

J.M. Ortega & W.C. Rheinboldt (1970): *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York.

A.M. Ostrowski (1966): *Solution of Equations and Systems of Equations*. Academic Press, New York, 2nd edition. [MA 65/27]

VI.1 Méthode des approximations successives

On considère le problème du calcul d'un *point fixe* de l'application $\Phi : I\!R^n \to I\!R^n$; c.-à-d., on cherche $x \in I\!R^n$ tel que

$$x = \Phi(x). \tag{1.1}$$

Les problèmes (0.1) et (1.1) sont équivalents et il y a beaucoup de possibilités pour écrire (0.1) sous la forme (1.1). Par exemple, on peut définir Φ comme $\Phi(x) = x - f(x)$ ou $\Phi(x) = x - Bf(x)$ (ici B est soit un nombre non-nul, soit une matrice bien choisie).

Pour résoudre (1.1), on se donne une approximation initiale x_0 (arbitraire) et on considère la méthode itérative

$$x_{k+1} = \Phi(x_k). \tag{1.2}$$

Si la suite $\{x_k\}$ converge, disons vers $a \in I\!R^n$, et si la fonction $\Phi(x)$ est continue en a, la limite a est une solution de (1.1). Mais que peut-on dire sur l'erreur?

Théorème 1.1 *Si $\Phi(x)$ est 2 fois continûment différentiable et si $a \in I\!R^n$ est une solution de (1.1), l'erreur $e_k = x_k - a$ satisfait*

$$e_{k+1} = \Phi'(a)\, e_k + \mathcal{O}(\|e_k\|^2). \tag{1.3}$$

Démonstration. Comme $a = \Phi(a)$, on obtient

$$e_{k+1} = x_{k+1} - a = \Phi(x_k) - \Phi(a) = \Phi'(a)e_k + \mathcal{O}(\|e_k\|^2). \qquad \square$$

On peut tirer plusieurs conclusions de la formule (1.3):

- si $\Phi'(a)$ possède une valeur propre λ_1 satisfaisant $|\lambda_1| > 1$, la composante de e_k dans la direction du vecteur propre v_1 va être agrandie – l'itération *ne converge pas* vers a.
- si toutes les valeurs propres de $\Phi'(a)$ satisfont $|\lambda_i| < 1$, on peut choisir une norme dans $I\!R^n$ telle que pour la norme matricielle correspondante $\|\Phi'(a)\| < 1$. Ceci et (1.3) impliquent que, pour $\|e_k\|$ suffisamment petit, on a $\|e_{k+1}\| \leq \alpha\|e_k\|$ où α est un nombre entre $\|\Phi'(a)\|$ et 1. L'erreur e_k converge donc vers zéro.

Exemple. Pour résoudre numériquement $y' = f(y)$, on peut appliquer la méthode d'Euler implicite

$$y_1 = y_0 + hf(y_1) \tag{1.4}$$

qui définit implicitement l'approximation y_1 de la solution après un pas de longueur h. L'équation (1.4) est déjà sous la forme (1.1) avec $\Phi(x) = y_0 + hf(x)$. Si h est suffisamment petit, les valeurs propres de $\Phi'(y_1) = hf'(y_1)$ sont petites et l'itération (1.2) converge.

Critère pour arrêter l'itération. En pratique, on s'intéresse à une approximation x_k qui satisfasse $\|x_k - a\| \leq$ *tol*. Une possibilité est d'accepter x_k comme approximation de la solution dès que $\|x_k - x_{k-1}\| \leq$ *tol*.

Le critère qui va suivre est basé sur l'hypothèse que Φ soit une contraction et que

$$\|x_k - x_{k-1}\| \leq \alpha\|x_{k-1} - x_{k-2}\| \qquad \text{avec} \qquad \alpha < 1.$$

En appliquant l'inégalité du triangle à l'identité

$$x_k - a = (x_k - x_{k+1}) + (x_{k+1} - x_{k+2}) + \ldots$$

(en cas de convergence), on obtient

$$\|x_k - a\| \leq \frac{\alpha}{1-\alpha}\|x_k - x_{k-1}\|. \tag{1.5}$$

Le facteur de contractivité peut être estimé par

$$\alpha_k = \|x_k - x_{k-1}\| / \|x_{k-1} - x_{k-2}\|, \qquad k \geq 2.$$

L'idée est d'arrêter l'itération quand

$$\frac{\alpha_k}{1-\alpha_k}\|x_k - x_{k-1}\| \leq \textit{tol} \tag{1.6}$$

et d'accepter x_k comme approximation de la solution.

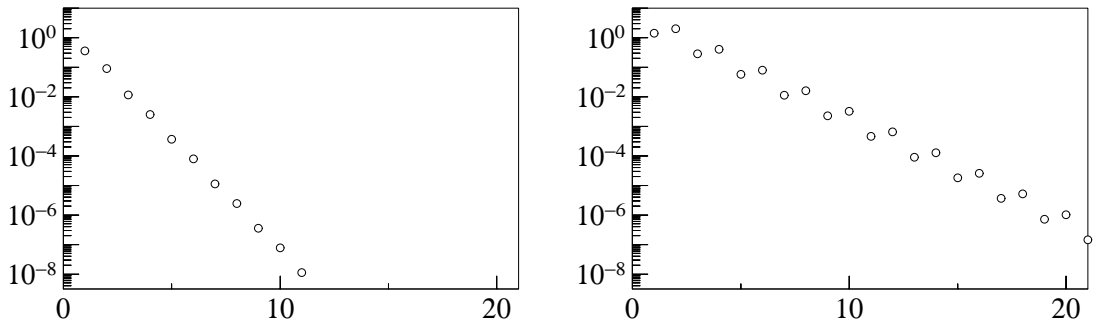

FIG. VI.1: Convergence des itérations (1.7)

Exemples. Pour les deux itérations

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.3 \begin{pmatrix} y_k \\ -\sin x_k \end{pmatrix}, \qquad \begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} 0 & -0.1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix} \tag{1.7}$$

la norme euclidienne de l'erreur de (x_k, y_k) est dessinée dans la figure VI.1 (à gauche pour la première itération et à droite pour la deuxième). On peut bien observer la convergence linéaire. Le rayon spectral de $\Phi'(a)$ vaut $\rho \approx 0.177$ et $\rho \approx 0.447$ respectivement. C'est la raison pour laquelle la première itération converge plus rapidement que la seconde. Le dessin à droite montre aussi que la convergence n'est pas nécessairement monotone (pour la norme $\|\cdot\|_2$). Donc, le coefficient α_k de (1.6) peut être plus grand que 1 même si l'itération converge.

VI.2 Méthodes itératives pour systèmes linéaires

Pour la résolution des systèmes linéaires

$$Ax = b, \tag{2.1}$$

il y a des situations où les méthodes itératives sont très utiles. Par exemple, si la matrice A possède une très grande dimension et si beaucoup d'éléments de A sont nuls (matrice creuse), ce qui est le cas pour les discrétisations des équations aux dérivées partielles.

Pour se ramener à un problème de point fixe, on considère une décomposition $A = M - N$ ("splitting") et on définit l'itération

$$Mx_{k+1} = Nx_k + b, \qquad A = M - N. \tag{2.2}$$

Le choix de la décomposition $A = M - N$ est important pour la performance de la méthode. D'une part, M doit être choisie telle que le système (2.2) soit beaucoup plus facile à résoudre que le système (2.1). D'autre part, les valeurs propres de la matrice $M^{-1}N$ doivent satisfaire $|\lambda_i| < 1$ pour que l'itération (2.2) converge.

Il y a beaucoup de possibilités de définir la décomposition $A = M - N$. Si l'on dénote

$$L = \begin{pmatrix} 0 & & & \\ a_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ a_{n1} & \dots & a_{n,n-1} & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_{n-1,n} \\ & & & 0 \end{pmatrix}$$

et $D = \operatorname{diag}(a_{11}, \dots, a_{nn})$ (afin que $A = L + D + U$) les itérations les plus connues sont:

- *Jacobi* : $M = D$, $N = -L - U$;
- *Gauss-Seidel* : $M = D + L$, $N = -U$.

Pour ces deux méthodes, la matrice M est choisie de manière à ce que le système (2.2) soit très facile à résoudre. Un avantage de la méthode de Gauss-Seidel est le fait que pour le calcul de la composante x_i^{k+1} du vecteur x_{k+1} on n'a besoin que des composantes $x_{i+1}^k, \ldots, x_n^k$ du vecteur x_k. Alors, on peut utiliser la même variable pour x_i^{k+1} que pour x_i^k. Une itération devient donc simplement:

for $i = 1, \ldots, n$
$x_i = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j - \sum_{j=i+1}^{n} a_{ij}x_j)/a_{ii}$.

Une modification intéressante de cet algorithme est la *méthode SOR* ("successive over-relaxation"):

$$\begin{aligned} x_{k+1} &= (1-\omega)x_k + \omega\widehat{x} \\ D\widehat{x} &= b - Lx_{k+1} - Ux_k \end{aligned} \tag{2.3}$$

où ω est un paramètre donné (pour $\omega = 1$, SOR se réduit à Gauss-Seidel). Elle est aussi simple à programmer que la précédente.

Les résultats suivants démontrent la convergence dans quelques situations particulières.

Théorème 2.1 *Si la matrice A est "diagonale dominante", c.-à-d.*

$$|a_{ii}| > \sum_{j\neq i} |a_{ij}| \qquad \textit{pour} \quad i = 1, \ldots, n, \tag{2.4}$$

alors l'itération de Jacobi converge.

Démonstration. Pour la méthode de Jacobi, on a $M^{-1}N = -D^{-1}(L+U)$. La condition (2.4) implique que $\|M^{-1}N\|_\infty < 1$ (voir la formule (4.5) du chapitre IV). □

Pour étudier la convergence de la méthode SOR (en particulier pour Gauss-Seidel), on l'écrit sous la forme équivalente

$$(D + \omega L)x_{k+1} = ((1-\omega)D - \omega U)x_k + \omega b$$

ou encore $x_{k+1} = H(\omega)x_k + \omega(D+\omega L)^{-1}b$ avec

$$H(\omega) = (D+\omega L)^{-1}((1-\omega)D - \omega U). \tag{2.5}$$

Théorème 2.2 *Si A est symétrique et définie positive et si $\omega \in (0,2)$, l'itération SOR, définie par (2.3), converge.*

Démonstration. Il faut démontrer que toutes les valeurs propres de $H(\omega)$ satisfont $|\lambda| < 1$. Soient λ une valeur propre et $x \neq 0$ le vecteur propre correspondant:

$$((1-\omega)D - \omega U)x = \lambda(D+\omega L)x. \tag{2.6}$$

Comme $(1-\omega)D - \omega U = D - \omega A + \omega L$, la formule (2.6) devient

$$\omega Ax = (1-\lambda)(D+\omega L)x. \tag{2.7}$$

En multipliant (2.6) et (2.7) avec x^*, on obtient pour $\alpha := x^*(D+\omega L)x$ (observer que $U = L^T$)

$$\lambda\alpha + \overline{\alpha} = (2-\omega)x^*Dx > 0, \qquad (1-\lambda)\alpha = \omega x^*Ax =: \beta > 0. \tag{2.8}$$

On en déduit

$$0 < \lambda\alpha + \overline{\alpha} = \beta\Big(\frac{\lambda}{1-\lambda} + \frac{1}{1-\overline{\lambda}}\Big) = \beta \cdot \frac{1-|\lambda|^2}{|1-\lambda|^2},$$

ce qui implique $|\lambda| < 1$. □

Pour quelques matrices, on connaît la valeur de ω qui minimise le rayon spectral de $H(\omega)$ et, par conséquent, qui accélère la convergence par rapport à l'itération de Gauss-Seidel. D'autres méthodes itératives pour des systèmes linéaires sont SSOR ("symmetric successive over-relaxation") et la méthode du gradient conjugué (avec préconditionnement). Voir le livre de Golub & Van Loan, déjà mentionné au chapitre IV, et les livres classiques

L.A. Hageman & D.M. Young (1981): *Applied Iterative Methods*. Academic Press, New York.

R.S. Varga (1962): *Matrix Iterative Methods*. Prentice Hall, Englewood Cliffs, New Jersey.

VI.3 Méthode de Newton

Considérons le problème de la résolution d'un système d'équations non linéaires

$$f(x) = 0 \tag{3.1}$$

où la fonction $f : I\!R^n \to I\!R^n$ est supposée être au moins une fois différentiable. Si $x_0 \in I\!R^n$ est une approximation de la solution cherchée, on linéarise $f(x)$ autour de x_0

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

et on calcule le zéro de cette linéarisation. Si l'on répète cette procédure avec la nouvelle approximation, on obtient l'algorithme suivant:

for $k = 0, 1, 2, \ldots$
  calculer $f(x_k)$ et $f'(x_k)$
  $f'(x_k)\Delta x_k = -f(x_k)$ (système linéaire à résoudre)
  $x_{k+1} = x_k + \Delta x_k$
end for

Exemple 3.1 La méthode d'Euler implicite (voir (1.4)), appliquée à l'équation différentielle $x' = y$, $y' = 10(1 - x^2)y - x$ avec pour valeurs initiales $\xi = 2$, $\eta = -0.66$, nous conduit à l'équation non linéaire

$$\begin{aligned} x &= \xi + h \cdot y \\ y &= \eta + h \cdot (10(1 - x^2)y - x). \end{aligned} \tag{3.2}$$

L'itération (1.2) ne converge que pour h suffisamment petit. Par exemple, pour $h = 0.3$ elle diverge et on est obligé d'utiliser un autre algorithme. La méthode de Newton

$$\begin{pmatrix} 1 & -h \\ h(20x_k y_k + 1) & 1 - 10h(1 - x_k^2) \end{pmatrix} \begin{pmatrix} x_{k+1} - x_k \\ y_{k+1} - y_k \end{pmatrix} = - \begin{pmatrix} x_k - \xi - hy_k \\ y_k - \eta - h(10(1 - x_k^2)y_k - x_k) \end{pmatrix}$$

converge sans difficulté avec $h = 0.3$ si l'on commence l'itération avec $x_0 = \xi$ et $y_0 = \eta$ (voir le tableau VI.1 pour les valeurs de x_k, y_k et les erreurs, mesurées dans la norme euclidienne).

TAB. VI.1: Convergence de la méthode de Newton

k	x_k	y_k	erreur
0	2.00000000000000	−0.660000000000000	$5.31 \cdot 10^{-1}$
1	1.95099818511797	−0.163339382940109	$3.38 \cdot 10^{-2}$
2	1.96084279415163	−0.130524019494582	$4.27 \cdot 10^{-4}$
3	1.96072023704926	−0.130932543169149	$6.65 \cdot 10^{-8}$
4	1.96072021795300	−0.130932606823320	$1.79 \cdot 10^{-15}$

Pour étudier la convergence de la méthode de Newton, considérons un terme de plus dans la série de Taylor:

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k, x - x_k) + \mathcal{O}(\|x - x_k\|^3). \tag{3.3}$$

Si l'on pose $x = a$ dans cette formule (avec a comme solution de (3.1)) et si l'on soustrait

$$0 = f(x_k) + f'(x_k)(x_{k+1} - x_k)$$

(définition de x_{k+1}), on obtient pour l'erreur $e_k = x_k - a$ la formule

$$0 = f'(x_k)(-e_{k+1}) + \frac{1}{2}f''(x_k)(e_k, e_k) + \mathcal{O}(\|e_k\|^3).$$

On vient de démontrer le résultat suivant:

Théorème 3.2 *Supposons que $f(x)$ soit 3 fois continûment différentiable et que $f'(x)$ soit inversible dans un voisinage de a (solution de (3.1)). Alors, pour x_k suffisamment proche de a, l'erreur $e_k = x_k - a$ de la méthode de Newton satisfait*

$$e_{k+1} = \frac{1}{2}(f'(x_k))^{-1}f''(x_k)(e_k, e_k) + \mathcal{O}(\|e_k\|^3).) \tag{3.4}$$

Donc, la convergence de cette méthode est quadratique. □

Ce théorème montre la convergence *locale* de la méthode de Newton, c.-à-d., si x_0 est proche d'une solution a de (3.1), la suite $\{x_k\}$ converge vers a. Concernant la convergence *globale*, on en sait très peu et on ne sait analyser seulement que quelques cas de fonctions simples. L'exemple le plus connu est

$$f(z) = z^3 - 1 \qquad \text{ou} \qquad f(x, y) = \begin{pmatrix} x^3 - 3xy^2 - 1 \\ 3x^2y - y^3 \end{pmatrix} \tag{3.5}$$

($z = x + iy$) pour lequel l'itération devient

$$z_{k+1} = z_k - \frac{z_k^3 - 1}{3z_k^2} = \frac{1}{3}\Big(2z_k + \frac{1}{z_k^2}\Big). \tag{3.6}$$

Il est intéressant de déterminer les ensembles (bassins d'attraction)

$$A(a) = \{z_0 \in \mathbb{C} \mid \{z_k\} \quad \text{converge vers } a\} \tag{3.7}$$

pour les trois solutions 1, $(-1 \pm i\sqrt{3})/2$ de $f(z) = 0$. Un calcul par ordinateur donne la figure VI.2. Les z_0 du domaine blanc entraînent une convergence vers $a = 1$, ceux du domaine gris vers $a = (-1 - i\sqrt{3})/2$ et ceux du domaine noir vers $a = (-1 + i\sqrt{3})/2$. On observe que la suite $\{z_k\}$ ne converge pas nécessairement vers la solution la plus proche de z_0.

Calcul numérique de la matrice jacobienne

En pratique, il arrive souvent que la forme analytique de la matrice $f'(x)$ est inconnue. Dans cette situation on approche les éléments $\partial f_i/\partial x_j$ de la matrice jacobienne par

$$\frac{\partial f_i}{\partial x_j}(x_1, \ldots, x_n) \approx \frac{f_i(x_1, \ldots, x_j + \delta, \ldots, x_n) - f_i(x_1, \ldots, x_n)}{\delta}. \tag{3.8}$$

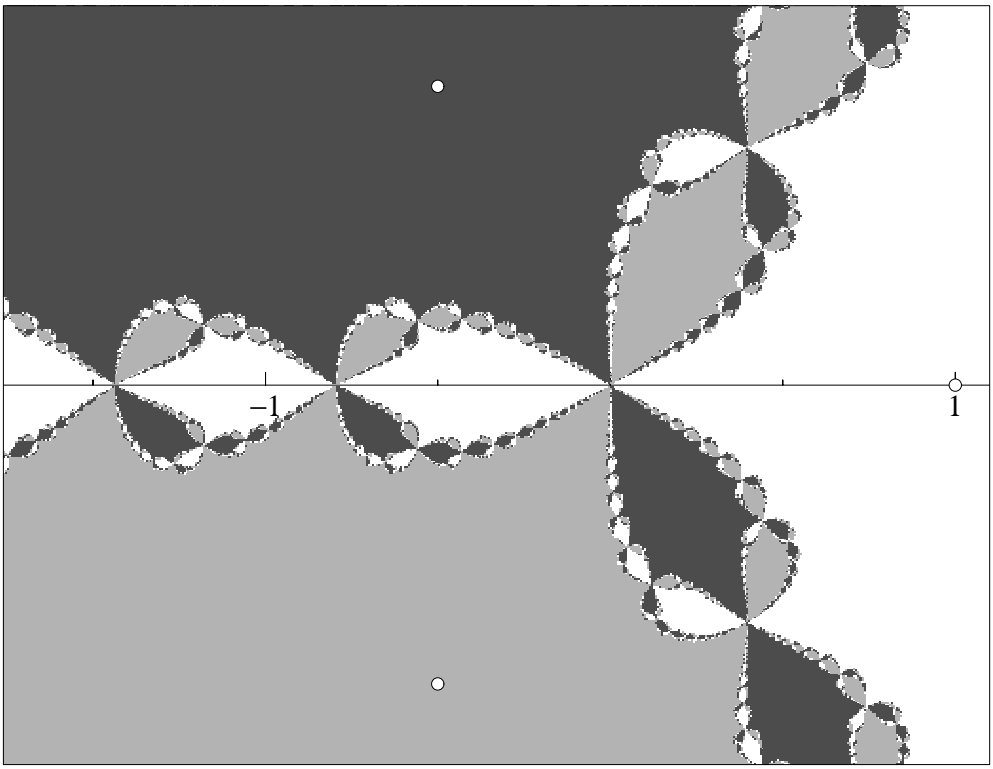

FIG. VI.2: Bassins d'attraction pour l'itération (3.6)

Mais comment doit-on choisir le δ? Pour simplifier la notation, considérons une fonction à une variable et notons-la par $g(x)$.

Un développement en série de Taylor montre que

$$\frac{g(x+\delta)-g(x)}{\delta} = g'(x) + \frac{\delta}{2}g''(x) + \mathcal{O}(\delta^2). \tag{3.9}$$

En conséquence, δ doit être petit pour que *l'erreur de la discrétisation* ne soit pas trop grande. D'autre part, la soustraction dans (3.9) est très mal conditionnée. Une étude des *erreurs d'arrondi* montre que l'expression obtenue par un calcul en virgule flottante vaut

$$\begin{aligned}&\frac{g((x+\delta)(1+\epsilon_1))(1+\epsilon_2)-g(x)(1+\epsilon_3)}{\delta}\\ &\quad\approx \frac{g(x+\delta)-g(x)}{\delta} + \frac{1}{\delta}\Big(g'(x+\delta)(x+\delta)\epsilon_1 + g(x+\delta)\epsilon_2 - g(x)\epsilon_3\Big)\end{aligned}$$

où $|\epsilon_i| \le$ *eps* (*eps* est la précision de l'ordinateur, voir section II.3). L'idée est de choisir δ afin que les deux erreurs soient de la même grandeur :

$$\frac{\delta}{2}\big(\ldots\big) \approx \frac{1}{\delta}\big(\ldots\big)\mathit{eps}. \tag{3.10}$$

Donc, on prend par exemple

$$\delta = \sqrt{\mathit{eps}} \qquad \text{ou} \qquad \delta = \sqrt{\mathit{eps}(1+|x|)}. \tag{3.11}$$

Modifications de la méthode de Newton

Si la dimension du problème (3.1) est très grande et/ou l'évaluation de la matrice $f'(x)$ est très coûteuse, on peut remplacer la définition de Δx_k par

$$f'(x_0)\Delta x_k = -f(x_k). \tag{3.12}$$

Dans ce cas, il suffit de calculer une fois pour toutes la décomposition LR de $f'(x_0)$, ce qui facilite grandement la résolution des systèmes linéaires successifs. Mais on perd la convergence quadratique car (3.12) n'est rien d'autre qu'une itération (1.2) avec $\Phi(x) = x - (f'(x_0))^{-1}f(x)$ et $\Phi'(a)$ est non nul en général.

En cas de mauvaise convergence, on remplace la définition de x_{k+1} par

$$x_{k+1} = x_k + \lambda_k \Delta x_k \tag{3.13}$$

et on détermine λ_k de façon à ce que $\|f(x_k + \lambda \Delta x_k)\|$ soit minimale.

VI.4 Méthode de Gauss-Newton

Dans ce paragraphe, on va s'intéresser à des systèmes non linéaires surdéterminés. On considère une fonction $f : I\!R^n \to I\!R^m$ où $m > n$ et on cherche une solution de $f(x) = 0$. Evidemment, comme on a plus de conditions que d'inconnues, ce problème ne possède en général pas de solution. On se contente donc de trouver un vecteur $x \in I\!R^n$ tel que

$$\|f(x)\|_2 \to \min . \tag{4.1}$$

Si $f(x)$ est différentiable, la fonction $F(x) = \|f(x)\|_2^2$ est aussi différentiable et une condition nécessaire pour que x soit un minimum local est d'avoir $F'(x) = 0$, c.-à-d.

$$f'(x)^T f(x) = 0. \tag{4.2}$$

Une possibilité pour résoudre (4.1) est d'appliquer une méthode itérative, par exemple la méthode de Newton, au système (4.2). Ceci nécessite le calcul de la deuxième dérivée de $f(x)$.

Une autre possibilité est de linéariser $f(x)$ dans (4.1) autour d'une approximation x_0 de la solution et de calculer x_1 de

$$\|f(x_0) + f'(x_0)(x_1 - x_0)\|_2 \to \min . \tag{4.3}$$

Une répétition de cette idée donne l'algorithme suivant (méthode de Gauss-Newton)

for $k = 0, 1, 2, \ldots$
  calculer $f(x_k)$ et $f'(x_k)$
  déterminer Δx_k de $\|f(x_k) + f'(x_k)\Delta x_k\|_2 \to \min$ (moindres carrés)
  $x_{k+1} = x_k + \Delta x_k$
end for

Pour calculer Δx_k on peut soit résoudre les équations normales (section IV.6)

$$(f'(x_k))^T f'(x_k)\Delta x_k = -(f'(x_k))^T f(x_k) \tag{4.4}$$

soit calculer la décomposition QR de $f'(x_k)$ et appliquer l'algorithme de la section IV.7.

Etude de la convergence

Pour étudier la convergence de la méthode de Gauss-Newton, développons la fonction

$$g(x) = (f'(x))^T f(x), \qquad g_i(x) = \sum_{j=1}^{m} \frac{\partial f_j}{\partial x_i}(x)\, f_j(x) \tag{4.5}$$

en série de Taylor. Pour ceci, calculons

$$\frac{\partial g_i}{\partial x_\ell} = \sum_{j=1}^{m} \frac{\partial^2 f_j}{\partial x_\ell \partial x_i}(x)\, f_j(x) + \sum_{j=1}^{m} \frac{\partial f_j}{\partial x_i}(x)\, \frac{\partial f_j}{\partial x_\ell}(x). \tag{4.6}$$

Matriciellement, la formule (4.6) s'écrit

$$g'(x) = B(x)(f(x),\ \cdot\ ) + (f'(x))^T f'(x) \tag{4.7}$$

où $B(x) : I\!R^m \times I\!R^n \to I\!R^n$ est l'application bilinéaire définie par

$$\Big(B(x)(u,v)\Big)_i = \sum_{\ell=1}^{n}\sum_{j=1}^{m} \frac{\partial^2 f_j}{\partial x_\ell \partial x_i}(x) \cdot u_j \cdot v_\ell.$$

Avec cette notation, la formule de Taylor donne

$$g(x) = (f'(x_k))^T f(x_k) + (f'(x_k))^T f'(x_k)(x-x_k) + B(x_k)(f(x_k), x-x_k) + \mathcal{O}(\|x-x_k\|^2). \tag{4.8}$$

Soit maintenant a une solution de (4.1). Elle satisfait $g(a) = 0$. En posant $x = a$ dans (4.8) et en soustrayant (4.4), nous obtenons ainsi le résultat suivant.

Théorème 4.1 *Supposons que $f : I\!R^n \to I\!R^m$ (avec $m > n$) soit 3 fois continûment différentiable, que le rang de $f'(x)$ soit maximal et que a soit une solution de (4.1). Alors, pour x_k suffisamment proche de a, l'erreur $e_k = x_k - a$ de la méthode de Gauss-Newton satisfait*

$$e_{k+1} = -\Big((f'(x_k))^T f'(x_k)\Big)^{-1} B(x_k)(f(x_k), e_k) + \mathcal{O}(\|e_k\|^2). \qquad \square$$

Une conséquence de cette formule est :

- en général, la convergence est linéaire;
- on a convergence quadratique si $f(a) = 0$;
- si $f(a)$ est trop grand, la méthode diverge.

Terminons ce chapitre avec une application typique de cet algorithme.

Exemple (Identification de paramètres). La figure VI.3 montre une photographie de la Vallée Blanche (prise par G. Wanner). On y reconnaît le Col des Grandes Jorasses, l'Aiguille du Géant, l'Aiguille Blanche de Peterey, l'Aiguille du Tacul, le Petit Rognon et l'Aiguille du Moine. La figure VI.4 est une copie d'une carte géographique de cette région. Le problème consiste à trouver la position de la caméra, ses caractéristiques (foyer) et les angles d'inclinaison.

Pour la formulation mathématique de ce problème, nous avons choisi des coordonnées (u, v) sur la photographie (figure VI.3, l'origine est au centre) et des coordonnées (x, y, z) sur la carte (z représente l'altitude). Les valeurs mesurées pour les 6 points reconnus sont données dans le tableau VI.2.

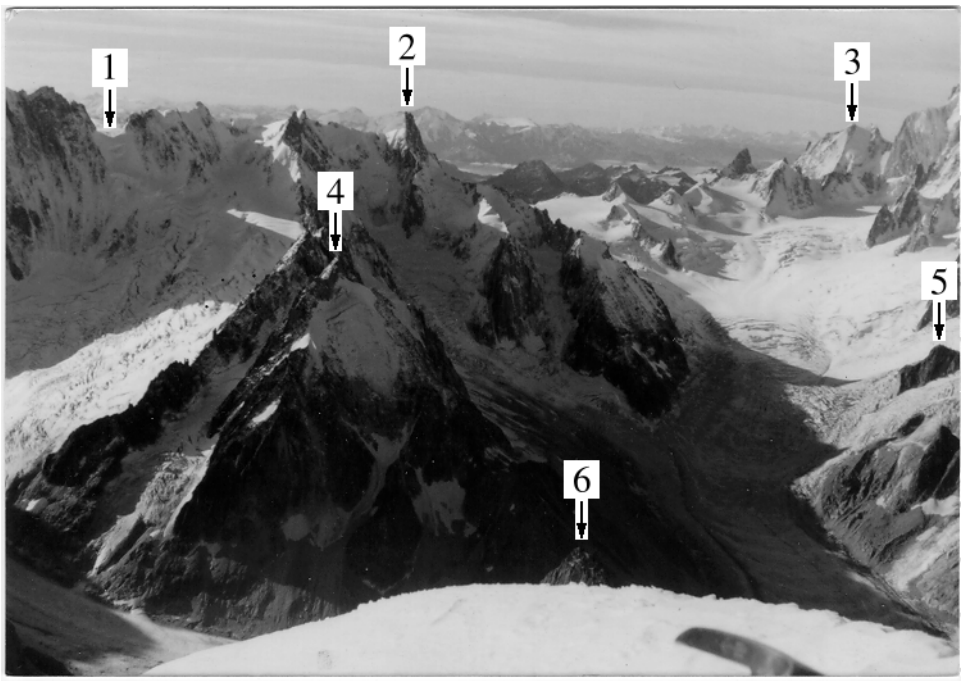

FIG. VI.3: Photographie de la Vallée Blanche

TAB. VI.2: Les données pour le problème “Vallée Blanche”

k	u_k	v_k	x_k	y_k	z_k
1. Col des Grandes Jorasses	-0.0480	0.0290	9855	5680	3825
2. Aiguille du Géant	-0.0100	0.0305	8170	5020	4013
3. Aig. Blanche de Peterey	0.0490	0.0285	2885	730	4107
4. Aiguille de Tacul	-0.0190	0.0115	8900	7530	3444
5. Petit Rognon	0.0600	-0.0005	5700	7025	3008
6. Aiguille du Moine	0.0125	-0.0270	8980	11120	3412

Pour fixer les *inconnues*, notons par $(\widehat{x}, \widehat{y}, \widehat{z})$ la position du foyer de la caméra, par le vecteur (a, b, c) la direction de vue avec norme correspondant à la distance du plan du film et par θ l'angle de rotation de la caméra autour du vecteur (a, b, c). On a donc 7 paramètres à trouver. De plus, considérons la base orthogonale

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \qquad h = \frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} b \\ -a \\ 0 \end{pmatrix}, \qquad g = \frac{1}{\sqrt{(a^2+b^2)(a^2+b^2+c^2)}} \begin{pmatrix} -ac \\ -bc \\ a^2+b^2 \end{pmatrix} \tag{4.9}$$

dont les deux vecteurs h et g engendrent le plan du film, h étant horizontal et g vertical. Alors, le vecteur dans l'espace qui montre du centre $(\widehat{x}, \widehat{y}, \widehat{z})$ de la lentille vers le point (u_k, v_k) sur le film est donné par

$$\begin{pmatrix} w_{1k} \\ w_{2k} \\ w_{3k} \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \alpha_k \cdot h + \beta_k \cdot g \qquad \text{où} \qquad \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u_k \\ v_k \end{pmatrix}.$$

Les conditions à satisfaire sont que les vecteurs (w_{1k}, w_{2k}, w_{3k}) et $(x_k - \widehat{x}, y_k - \widehat{y}, z_k - \widehat{z})$ soient

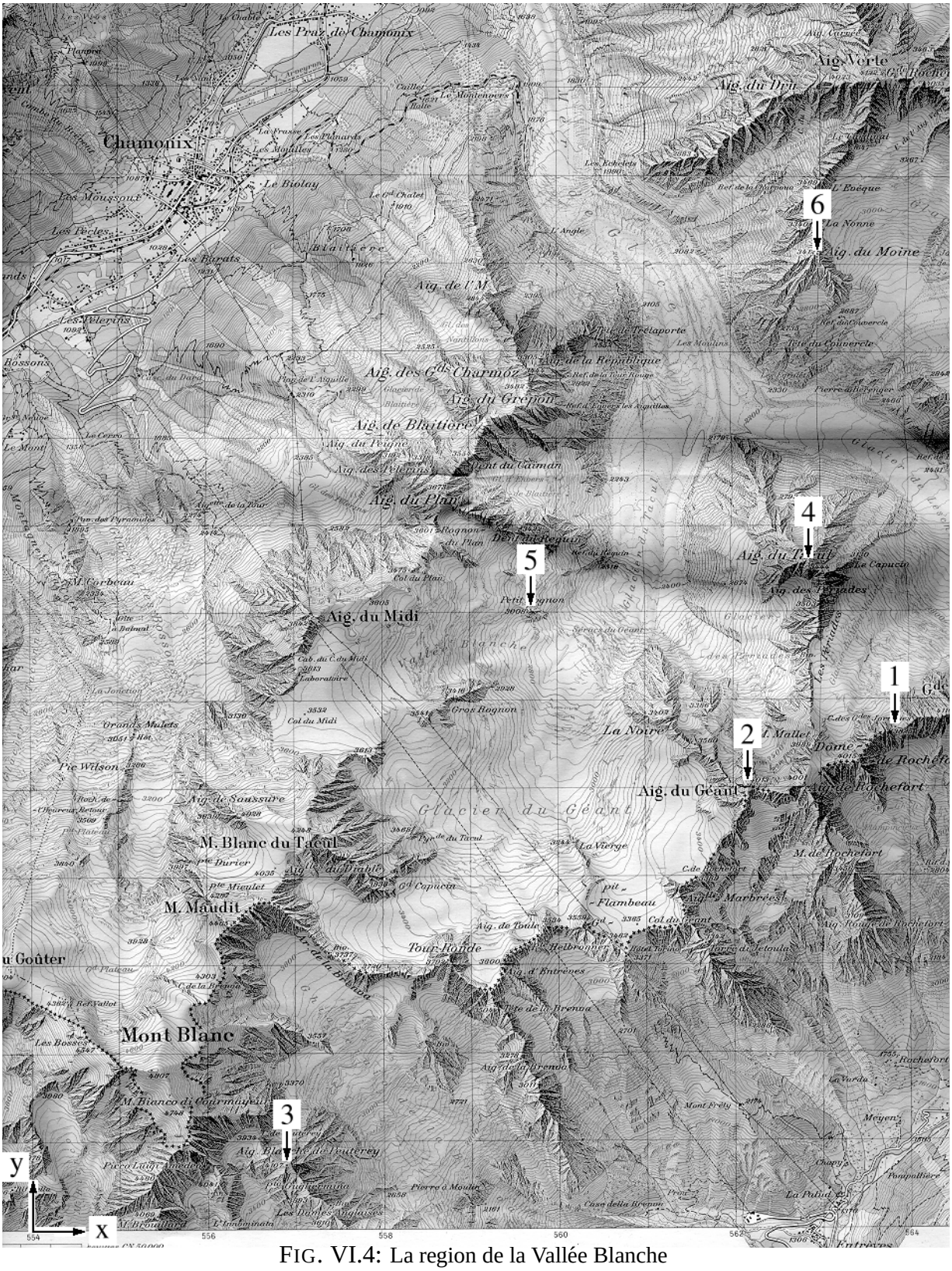

FIG. VI.4: La region de la Vallée Blanche

colinéaires. Ceci donne deux conditions pour chaque k. Mais, par symétrie, il est plus naturel de considérer les trois conditions (pour chaque k)

$$0 = \begin{pmatrix} w_{1k} \\ w_{2k} \\ w_{3k} \end{pmatrix} \times \begin{pmatrix} x_k - \widehat{x} \\ y_k - \widehat{y} \\ z_k - \widehat{z} \end{pmatrix} = \begin{pmatrix} w_{2k}(z_k - \widehat{z}) - w_{3k}(y_k - \widehat{y}) \\ w_{3k}(x_k - \widehat{x}) - w_{1k}(z_k - \widehat{z}) \\ w_{1k}(y_k - \widehat{y}) - w_{2k}(x_k - \widehat{x}) \end{pmatrix}. \tag{4.10}$$

En tout, ceci donne $3 \cdot 6 = 18$ conditions pour 7 inconnues. Voici le programme FORTRAN pour la fonction $f : I\!R^7 \to I\!R^{18}$.

```
      SUBROUTINE FCN(N,XSOL,M,F)
      IMPLICIT REAL*8 (A-H,O-Z)
      PARAMETER(ND=50)
      DIMENSION XSOL(N),F(M)
      COMMON/DAT/NDAT,XDAT(ND),YDAT(ND),ZDAT(ND),UDAT(ND),VDAT(ND)
C ------ NOTATION LOCALE --------
      XO=XSOL(1)
      YO=XSOL(2)
      ZO=XSOL(3)
      A=XSOL(4)
      B=XSOL(5)
      C=XSOL(6)
      THETA=XSOL(7)
C ------ VECTEUR HORIZONTAL ------
      RAC=SQRT(A*A+B*B)
      H1=B/RAC
      H2=-A/RAC
      H3=0.
C ----- VECTEUR VERTICAL ---------
      RAC=SQRT((A*C)**2+(B*C)**2+(A**2+B**2)**2)
      G1=-A*C/RAC
      G2=-B*C/RAC
      G3=(A**2+B**2)/RAC
C ------- LES POINTS --------
      DO I=1,NDAT
C ------ ROTATION INITIALE ------
         U= UDAT(I)*COS(THETA)+VDAT(I)*SIN(THETA)
         V=-UDAT(I)*SIN(THETA)+VDAT(I)*COS(THETA)
C ------ LES VECTEURS A ALIGNER -------
         W1=A+H1*U+G1*V
         W2=B+H2*U+G2*V
         W3=C+H3*U+G3*V
         Q1=XDAT(I)-XO
         Q2=YDAT(I)-YO
         Q3=ZDAT(I)-ZO
C -------- LES FONCTIONS A ANNULER -----------
         F(3*(I-1)+1)=W1*Q2-W2*Q1
         F(3*(I-1)+2)=W2*Q3-W3*Q2
         F(3*(I-1)+3)=W3*Q1-W1*Q3
      END DO
      RETURN
      END
```

En appliquant la méthode de Gauss-Newton à ce système avec comme valeurs initiales $\widehat{x} = 8000$, $\widehat{y} = 15000$, $\widehat{z} = 1000$, $a = 0$, $b = -1$, $c = 0$, $\theta = 0$, après peu d'itérations on obtient la solution $\widehat{x} = 9664$, $\widehat{y} = 13115$, $\widehat{z} = 4116$ avec une précision de 5 chiffres (voir le tableau VI.3). On a donc bien trouvé la position de la caméra. C'est à l'Aiguille Verte (altitude 4122) que Gerhard a pris cette photo.

TAB. VI.3: Convergence de la méthode de Gauss-Newton

k	$\widehat{x}_k$	$\widehat{y}_k$	$\widehat{z}_k$	a	b	c	θ
0	8000	15000	1000	0.000	−1.000	0.000	0.000
1	8030	9339	1169	−0.003	−0.085	−0.003	0.047
2	8680	11163	4017	−0.014	−0.114	−0.021	0.017
3	9577	13034	3993	−0.040	−0.167	−0.032	−0.094
4	9660	13107	4116	−0.043	−0.169	−0.032	−0.074
5	9664	13115	4116	−0.043	−0.169	−0.032	−0.074
6	9664	13115	4116	−0.043	−0.169	−0.032	−0.074

VI.5 Exercices

1. Pour une fonction $f : \mathbf{R} \to \mathbf{R}$, considrons l'itration $(x_k, x_{k-1}) \to x_{k+1}$ dfinie par

$$0 = f(x_k) + \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}(x_{k+1} - x_k). \tag{5.1}$$

   (a) Donner une interprtation gomtrique.

   (b) Dmontrer que l'erreur $e_k = |x_k - \zeta|$, o ζ est une racine simple de $f(x) = 0$, satisfait

$$e_{k+1} \approx C e_k e_{k-1}, \quad \text{avec} \quad C = \left| \frac{f''(\zeta)}{2f'(\zeta)} \right|.$$

   (c) Dduire de (b) que

$$e_{k+1} \approx D e_k^p, \quad \text{avec} \quad p = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618.$$

   (d) Soit $f(x)$ un polynme, nous avons que le travail pour valuer $f(x)$ et $f'(x)$ est approximativement le mme. Supposons que le cot des oprations d'additions, de soustractions, de multiplications et de divisions sont ngligeables par rapport l'valuation de $f(x)$. Quelle mthode choisiriez-vous entre la mthode de Newton et la mthode de la scante (5.1) pour calculer une racine de $f(x)$ travail gal?

2. Soit $D \subset \mathbf{R}$ et $f : D \to \mathbf{R}^n$ continûment différentiable. Pour un $x_0 \in D$ supposons que $f(x_0) \neq 0$ et que $f'(x_0)$ soit inversible. Montrer que

$$p_0 = -f'(x_0)^{-1} f(x_0)$$

   est la seule direction ayant la propiété suivante:
   Pour toute matrice inversible A, il existe $\lambda_0 > 0$ tel que pour $0 < \lambda < \lambda_0$

$$\|Af(x_0 + \lambda p_0)\|_2 < \|Af(x_0)\|_2 \ .$$