

Chapitre 3

Les systèmes non linéaires

Dans ce chapitre, nous nous intéressons à la résolution de systèmes non linéaires issus de la modélisation en mécanique, chimie ou en dynamique des populations. C'est une étape clé dans la résolution de la plupart des problèmes en mathématiques appliquées tenant compte de phénomènes non linéaires. Le problème s'écrit sous la forme

$$\left\{ \begin{array}{l} \text{trouver } \bar{x} \in E, \\ \text{tel que } f(\bar{x}) = 0, \end{array} \right.$$

où E est un ensemble donné. Il s'agit tout d'abord de proposer un cadre mathématique adapté à ce type de problème non linéaire (existence et unicité de solution, localisation de la solution). Différentes méthodes numériques seront proposées et la notion de convergence des algorithmes sera abordée (convergence locale et globale) ainsi que l'ordre de convergence indiquant la vitesse à laquelle la solution numérique se rapproche de la solution exacte.

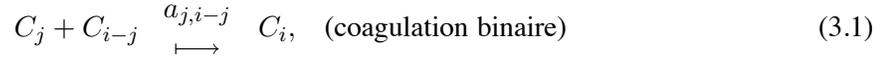
3.1 Introduction aux problèmes non linéaires

Nous commençons par un exemple simple de problème non linéaire.

3.1.1 Modèle de coagulation et fragmentation

Des modèles de coagulation et fragmentation ont été introduits par M. von Smoluchowski en 1916 et décrivent l'évolution en fonction du temps $t \geq 0$ de la densité $C_i(t)$ de particules de masse $i > 0$, où l'unité de masse est la masse d'une particule de référence (proton, soleil). Nous rencontrons ces modèles dans l'étude de la formation d'agrégats (polymères, étoiles). Ils décrivent les mécanismes par lesquels deux agrégats coagulent pour en former un plus gros tandis que dans le même temps d'autres particules se fragmentent en particules de plus petites masses ; la masse globale étant conservée au cours de chaque réaction de coalescence. Dans le cas le plus simple, les masses des agrégats ne prennent que des valeurs discrètes, l'évolution temporelle des densités des agrégats de masse i est donnée par un système d'équations différentielles fortement couplées. Nous nous intéressons ici non pas au processus d'évolution mais aux états d'équilibre. Plus précisément, C_i la densité de masse i

avec $i \in \mathbb{N}^*$ est le résultat d'une réaction de coagulation



et d'une réaction de fragmentation



où a et b désignent les taux de coagulation et de fragmentation, c'est-à-dire que $a_{i,j}$ représente la probabilité qu'une particule de masse i fusionne avec une particule de masse j pour donner une nouvelle particule de masse $i + j$ tandis que $b_{i,j}$ représente la probabilité qu'une particule de masse $i + j$ se scinde en deux particules de masse i et une autre de masse j . Nous supposons que les taux de coagulation et de fragmentation dépendent uniquement de la masse des particules. Nous obtenons alors à l'équilibre le modèle de coagulation-fragmentation suivant [28] : pour tout $i \in \mathbb{N}^*$

$$\begin{aligned} 0 &= \underbrace{\frac{1}{2} \sum_{j=1}^{i-1} a_{i-j,j} C_{i-j} C_j}_{C_i \text{ créée par coagulation}} - \underbrace{\sum_{j=1}^{\infty} a_{i,j} C_j C_i}_{C_i \text{ détruite après coagulation}} \\ &- \underbrace{\frac{1}{2} \sum_{j=1}^{i-1} b_{i-j,j} C_i}_{C_i \text{ détruite après fragmentation}} + \underbrace{\sum_{j=1}^{\infty} b_{i,j} C_{i+j}}_{C_i \text{ créée par fragmentation}}. \end{aligned}$$

La suite $(C_i)_{i \in \mathbb{N}^*}$ est donc solution d'un système infini d'équations non linéaires dont il est pratiquement impossible de calculer la solution exacte. Il est donc indispensable d'avoir recours à des algorithmes numériques qui réduisent ce problème en un système fini d'équations couplées ; dès lors une méthode numérique permet d'en calculer une solution approchée.

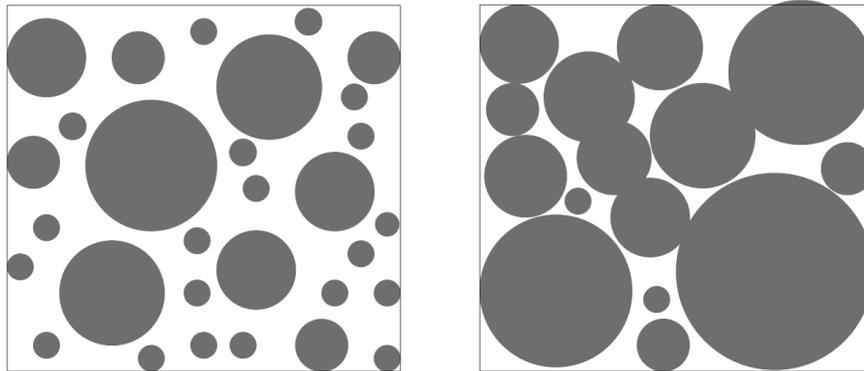


FIG. 3.1 – Exemple de l'évolution de la densité de particules $(C_i)_{i \in \mathbb{N}^*}$ obtenue en résolvant le modèle non linéaire de coagulation-fragmentation au temps $t = 0$ et $t = 1$.

3.1.2 Résultats généraux et définitions

Il est temps de formaliser ces problèmes non linéaires d'un point de vue mathématique. Considérons E un espace métrique où d est la distance associée¹ et F un \mathbb{R} -espace vectoriel, supposons que K soit un sous-ensemble de E , et f une fonction de K à valeurs dans F . Nous nous intéressons à la résolution du problème suivant

$$\begin{cases} \bar{x} \in K, \\ f(\bar{x}) = 0_F \end{cases} \quad (3.3)$$

et disons que \bar{x} est un zéro de la fonction f .

Ce problème est non linéaire dans le sens où la fonction f n'est plus seulement une fonction du type $f(x) = Ax - b$ avec $A \in \mathcal{M}_{n,n}(\mathbb{R})$, $b \in \mathbb{R}^n$ et $x \in \mathbb{R}^n$. Pour la résolution numérique de ce problème, nous ne pouvons donc pas utiliser les algorithmes développés dans la partie précédente mais nous verrons qu'il s'agit toujours de construire une suite d'approximations $(x^{(k)})_{k \in \mathbb{N}}$ dont la limite \bar{x} est solution de (3.3). Précisons d'abord la notion de convergence de la suite $(x^{(k)})_{k \in \mathbb{N}}$, puis introduisons la notion d'ordre de convergence.

Définition 3.1.1 Soient E un espace métrique muni de la distance $d(\cdot, \cdot)$ et $(x^{(k)})_{k \in \mathbb{N}}$ une suite de E approchant la solution \bar{x} de (3.3) et vérifiant $x^{(0)} = x_0 \in U$. Nous disons que la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers \bar{x} si $\lim_{k \rightarrow \infty} d(x^{(k)}, \bar{x}) = 0$. De plus, nous disons que la méthode itérative fournissant les valeurs de $(x^{(k)})_{k \in \mathbb{N}}$ est d'ordre p , s'il existe deux constantes C_1 et $C_2 > 0$ telles que

$$C_1 \leq \lim_{k \rightarrow \infty} \frac{d(x^{(k+1)}, \bar{x})}{d(x^{(k)}, \bar{x})^p} \leq C_2.$$

La notion de vitesse de convergence est évidemment essentielle en analyse numérique. Afin de diminuer le temps de calcul, nous choisissons souvent d'utiliser l'algorithme qui converge le plus rapidement, c'est-à-dire d'ordre le plus élevé. En effet, lorsque $x^{(k)}$ est suffisamment proche de \bar{x} pour la distance d , l'itéré suivant sera de l'ordre de $d(x^{(k+1)}, \bar{x}) \leq C_2 d(x^{(k)}, \bar{x})^p$ qui est d'autant plus petit que p est grand.

Une autre notion importante est la distinction entre la convergence *locale* et la convergence *globale*.

Définition 3.1.2 Soient E un espace métrique et $(x^{(k)})_{k \in \mathbb{N}} \subset E$ une suite d'approximation de la solution \bar{x} de (3.3) telle que $x^{(0)} = x_0$.

- Nous disons que la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge globalement vers $\bar{x} \in K$ si pour tout $x_0 \in K$, la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers \bar{x} .
- Nous disons que la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge localement vers $\bar{x} \in K$ s'il existe un voisinage $V_{\bar{x}}$ de \bar{x} , tel que pour tout $x_0 \in V_{\bar{x}}$, la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers \bar{x} .

Dans la pratique, nous préférons souvent des méthodes qui donnent la convergence globale de la solution à moins bien sûr d'avoir une idée *a priori* de la localisation de la solution.

Dans une première partie, nous présentons une méthode générale et simple, pour laquelle la recherche d'un zéro d'une fonction $f : K \subset E \rightarrow F$ est remplacée par la recherche d'un point

¹La distance d est une application de $E \times E$ à valeurs dans \mathbb{R}^+ telle que pour tout x, y et $z \in E$, $d(x, y) = d(y, x)$, $d(x, y) = 0 \Rightarrow x = y$ et $d(x, z) \leq d(x, y) + d(y, z)$.

fixe d'une nouvelle fonction Φ . Ensuite, nous proposons plusieurs algorithmes pour la résolution numérique de problèmes non linéaires dans le cas de fonctions à valeurs dans \mathbb{R} . Puis, nous traitons le cas $E = F = \mathbb{R}^n$, en proposant la méthode de Newton-Raphson pour les systèmes non linéaires.

3.2 Méthode de point fixe

Considérons E un espace métrique muni d'une distance d , K un sous-ensemble fermé de E et f une fonction : $K \subset E \rightarrow E$. La méthode de point fixe consiste à remplacer la fonction f par la fonction Φ en posant par exemple $\Phi(x) = f(x) + x$ tout en s'assurant que l'ensemble K est stable par l'application Φ , c'est-à-dire $\Phi(K) \subset K$, nous verrons qu'il y a plusieurs choix possibles pour cela. Nous remplaçons alors l'équation (3.3) par un nouveau problème : $\bar{x} \in K$, tel que $\Phi(\bar{x}) = \bar{x}$.

Présentons tout d'abord la méthode de Héron permettant d'approcher des racines carrées d'un nombre réel. Cet algorithme est basé sur le principe du point fixe. Proposons ensuite une méthode générale et indiquons les critères que doit satisfaire la fonction Φ pour assurer la convergence.

3.2.1 Méthode de Héron

La méthode de Héron² fournit un algorithme pour la recherche de la racine carrée d'un nombre réel positif a . Nous allons utiliser le fait que le réel $\alpha = \sqrt{a}$ n'est rien d'autre que la longueur du côté d'un carré dont l'aire est a . L'idée de base de la méthode de Héron va donc consister à construire un tel carré en construisant une suite de rectangles d'aire a qui converge vers le carré en question.

Considérons une valeur $x^{(0)}$ donnée, qui représente une estimation grossière de α . Choisissons alors le rectangle d'aire a ayant un côté de longueur $L = x^{(0)}$ et un second côté égal à $l = a/L = a/x^{(0)}$. Pour construire un nouveau rectangle qui soit plus proche du carré d'aire a , nous prenons simplement pour longueur d'un côté la moyenne des longueurs des côtés du rectangle précédent, c'est-à-dire, $L = x^{(1)} = (x^{(0)} + a/x^{(0)})/2$ tandis que le second côté aura pour longueur $l = a/L = a/x^{(1)}$. Nous continuons ce procédé jusqu'à ce que nous ne soyons plus en mesure de distinguer les longueurs des deux côtés, cela revient à construire une suite $(x^{(k)})_{k \geq 0}$ dont le terme général vérifie pour $k \in \mathbb{N}$, $x^{(k+1)} = (x^{(k)} + a/x^{(k)})/2$.

Par exemple, si nous recherchons la racine carrée de $a = 200$ nous obtenons alors les valeurs données dans le tableau ci-contre. Avec seulement quatre itérations, nous avons obtenu la racine cherchée à la précision machine.

k	$x^{(k)}$	$\frac{a}{x^{(k)}}$
0	10,000	20,000
1	15,000	13,333
2	14,166	14,117
3	14,142	14,142
4	14,142	14,142

²En référence à Héron d'Alexandrie qui vécut au premier siècle après J.-C. et fut un des grands mécaniciens de l'Antiquité. L'algorithme de calcul des racines carrées lui est attribué bien qu'il semble que cet algorithme fût déjà connu des Babyloniens.

Voyons maintenant comment généraliser cette approche pour la résolution d'une équation du type (3.3).

3.2.2 Méthode générale

Soit E un espace métrique et $K \subset E$ un sous-ensemble fermé de E . Comme nous l'avons déjà précisé, la méthode de point fixe consiste à d'abord remplacer l'équation (3.3) par une nouvelle équation

$$\begin{cases} \bar{x} \in K, \\ \\ \text{tel que } \Phi(\bar{x}) = \bar{x}, \end{cases} \quad (3.4)$$

où Φ est une fonction $\Phi : K \subset E \mapsto K$, nous pouvons choisir $\Phi(x) = f(x) + x$. Cependant, pour assurer qu'à chaque itération $x^{(k)} \in K$ puisque l'inconnue $\bar{x} \in K$, la fonction Φ doit satisfaire $\Phi(K) \subset K$ ce qui n'est pas facile à vérifier en général.

Par exemple, en reprenant la méthode de Héron, nous approchons la racine carrée de $a \in \mathbb{R}^+$, laquelle vérifie l'équation $f(x) := x^2 - a = 0$. Dans ce cas, nous avons pris la fonction Φ définie sur $K := \mathbb{R}^+$ et donnée par $\Phi(x) = (x + a/x)/2$, laquelle vérifie bien $\Phi(K) \subset K$ et $\Phi(\sqrt{a}) = \sqrt{a}$. Cependant nous aurions pu prendre aussi :

- $\Phi(x) = x^2 - a + x$, mais nous ne pouvons pas assurer la stabilité $\Phi(K) \subset K$ lorsque $K = \mathbb{R}^+$.
- $\Phi(x) = a/x$, et $K = \mathbb{R}^+$, cette fois nous avons bien $\Phi(\mathbb{R}^+) \subset \mathbb{R}^+$ mais nous ne pouvons pas toujours assurer la convergence de l'algorithme itératif.

En nous inspirant de la méthode de Héron, nous proposons alors un algorithme de point fixe pour approcher la solution de (3.3) dans le cas où f est une fonction de E dans E . Nous construisons une fonction Φ telle que la solution \bar{x} de (3.3) soit aussi un point fixe de Φ . Puis, pour $x_0 \in K \subset E$

$$\begin{cases} x^{(0)} = x_0, \\ \\ x^{(k+1)} = \Phi(x^{(k)}), \quad k \geq 0, \end{cases} \quad (3.5)$$

où Φ est définie de K dans K .

L'algorithme de point fixe général pour le problème (3.4) est alors donné, pour $\varepsilon > 0$ un petit paramètre et $x_0 \in K$, par

Algorithme 1. Algorithme de point fixe

Poser $x^{(0)} = x_0, \varepsilon^{(0)} = d(\Phi(x^{(0)}), x^{(0)})$ et $k = 0$

Tant que $\varepsilon^{(k)} \geq \varepsilon$

- calculer $x^{(k+1)} = \Phi(x^{(k)})$,

- calculer le résidu pour le test d'arrêt

$$\varepsilon^{(k+1)} := d(x^{(k+1)}, x^{(k)}),$$

- itérer $k \leftarrow k + 1$.

Fin de tant que.

Il reste à déterminer des critères sur la fonction Φ pour que la méthode de point fixe soit effectivement convergente.

Rappelons d'abord la définition d'une suite de Cauchy

Définition 3.2.1 Soient E un espace métrique et $(x^{(k)})_{k \in \mathbb{N}}$ une suite de E . Nous disons que $(x^{(k)})_{k \in \mathbb{N}}$ est une suite de Cauchy lorsqu'elle vérifie : pour tout $\varepsilon > 0$, il existe $N_\varepsilon > 0$ tel que pour tout $k, l \geq N_\varepsilon, d(x^{(k)}, x^{(l)}) \leq \varepsilon$.

Nous construisons une suite d'approximation $(x^{(k)})_{k \in \mathbb{N}}$ fournie par l'algorithme (3.5) et cherchons à établir la convergence de la suite $(x^{(k)})_{k \in \mathbb{N}}$ vers une solution \bar{x} de l'équation $f(\bar{x}) = 0$. Pour cela, nous donnons le théorème suivant.

Théorème 3.2.1 (Théorème du point fixe de Banach) Soit E un espace métrique muni de la distance d , complet c'est-à-dire que toutes les suites de Cauchy sont convergentes, $K \subset E$ un sous-ensemble fermé de E et $\Phi : K \mapsto K$ une fonction strictement contractante, c'est-à-dire qu'il existe une constante L telle que $0 < L < 1$ et

$$d(\Phi(x), \Phi(y)) \leq L d(x, y), \quad x, y \in K. \quad (3.6)$$

Alors

- il existe un unique point fixe $\bar{x} \in K$ tel que $\bar{x} = \Phi(\bar{x})$;
- pour toute donnée initiale $x_0 \in K$, la suite d'approximation définie par (3.5) converge vers le point fixe de Φ et vérifie $d(x^{(k+1)}, \bar{x}) \leq L d(x^{(k)}, \bar{x})$. Nous disons alors que la convergence est linéaire.

Démonstration : Soit $(x^{(k)})_{k \in \mathbb{N}}$ la suite définie par $x^{(k+1)} = \Phi(x^{(k)})$, pour $k \in \mathbb{N}$. Nous allons démontrer que c'est une suite de Cauchy. Tout d'abord, d'après l'hypothèse (3.6) nous avons

$$d(x^{(k+1)}, x^{(k)}) \leq L d(x^{(k)}, x^{(k-1)}) \leq \dots \leq L^k d(x^{(1)}, x^{(0)}).$$

Soient m et k deux entiers, nous obtenons alors

$$\begin{aligned} d(x^{(k+m)}, x^{(k)}) &\leq d(x^{(k+m)}, x^{(k+m-1)}) + d(x^{(k+m-1)}, x^{(k+m-2)}) + \dots \\ &\quad + d(x^{(k+1)}, x^{(k)}) \\ &\leq (L^{k+m-1} + \dots + L^k) d(x^{(1)}, x^{(0)}) \\ &\leq \frac{L^k}{1-L} d(x^{(1)}, x^{(0)}). \end{aligned}$$

Puisque L^k converge vers zéro, lorsque k tend vers l'infini, nous avons prouvé que $(x^{(k)})_{k \in \mathbb{N}}$ est une suite de Cauchy et donc puisque E est complet, le sous-ensemble fermé K l'est aussi et la suite $(x^{(k)})_{k \in \mathbb{N}}$ est convergente vers un point $\bar{x} \in K$. Signalons que la fonction Φ étant contractante, elle est continue donc en passant à la limite dans le schéma itératif, nous obtenons $\bar{x} = \Phi(\bar{x})$. Puis, enfin $d(\Phi(x^{(k)}), \Phi(\bar{x})) = d(x^{(k+1)}, \bar{x}) \leq L d(x^{(k)}, \bar{x})$.

Pour terminer la démonstration nous vérifions que le point fixe de l'application Φ est unique. En effet, soient \bar{x} et \bar{y} deux points fixes de Φ alors

$$0 \leq d(\bar{x}, \bar{y}) = d(\Phi(\bar{x}), \Phi(\bar{y})) \leq L d(\bar{x}, \bar{y}) < d(\bar{x}, \bar{y}),$$

d'où $0 \leq d(\bar{x}, \bar{y}) < d(\bar{x}, \bar{y})$, ce qui montre forcément que $\bar{x} = \bar{y}$. □

Corollaire 3.2.1 Soient $(E, \|\cdot\|)$ un \mathbb{R} -espace vectoriel normé complet (nous disons que c'est un espace de Banach), U un ouvert de E et $\Phi : U \mapsto U$ une fonction différentiable dont la différentielle est continue sur U . Supposons que $\bar{x} \in U \subset E$ soit un point fixe de Φ tel que $\|\Phi'(\bar{x})\| < 1$. Alors il existe δ tel que pour tout $x^{(0)} = x_0 \in \mathcal{B}(\bar{x}, \delta)$, boule de centre \bar{x} et de rayon $\delta > 0$, telle que la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers le point fixe \bar{x} . La convergence est dans ce cas seulement locale.

Notons bien que la méthode de point fixe n'est pas seulement une méthode numérique pour la résolution du problème non linéaire (3.3) puisqu'elle fournit également une démonstration de l'existence de solutions lorsque le problème est mis sous la forme (3.5). C'est d'ailleurs une méthode classique d'existence de solutions pour un problème non linéaire. Par la suite, nous donnons d'autres théorèmes de point fixe sous des hypothèses plus faibles.

Rappelons d'abord la notion de convexité qui joue un rôle important dans la résolution de problème non linéaire et d'optimisation.

Définition 3.2.2 Une partie K de E est dite convexe si, pour tout $x, y \in K$ et tout $\lambda \in \mathbb{R}$ vérifiant $0 \leq \lambda \leq 1$, le point $\lambda x + (1 - \lambda)y \in K$. Géométriquement, cela signifie que le segment de droite joignant deux points x et y de K est entièrement contenu dans K .

Définition 3.2.3 Nous disons qu'une partie K d'un espace vectoriel normé E est compacte si toute suite de K possède une suite extraite convergente.

Nous énonçons alors le théorème de Brouwer lorsque E est l'espace euclidien \mathbb{R}^n .

Théorème 3.2.2 (Théorème de point fixe de Brouwer) *Soient K un sous-ensemble convexe et compact³ de l'espace euclidien $E = \mathbb{R}^n$ et Φ une application continue de K dans K . Alors Φ possède un point fixe et lorsque $x_0 \in K$ la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers un $\bar{x} \in K$ tel que $\Phi(\bar{x}) = \bar{x}$.*

Nous pouvons également utiliser le Théorème de Schauder, démontré en 1930 qui est souvent utilisé pour traiter les problèmes d'existence d'équations aux dérivées partielles non linéaires et qui généralise le théorème précédent.

Théorème 3.2.3 (Théorème de point fixe de Schauder) *Soient E un \mathbb{R} -espace de Banach, $K \subset E$ une partie non vide, convexe, compacte de E et Φ une application continue de K dans K . Alors Φ possède un point fixe et pour tout $x_0 \in K$ la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers un $\bar{x} \in K$ tel que $\Phi(\bar{x}) = \bar{x}$.*

L'avantage de cette méthode est qu'elle est très générale lorsque la fonction va de E dans E et fonctionne en dimension quelconque sans faire intervenir aucune dérivée de la fonction Φ . Cependant, cette méthode n'est souvent pas très efficace puisque la convergence est seulement linéaire. Nous construisons par la suite des méthodes d'ordre plus élevé.

3.3 Vers la méthode de Newton-Raphson

Soit $[a, b] \subset \mathbb{R}$, nous considérons le cas simplifié de fonction f de \mathbb{R} dans \mathbb{R} ($E = F = \mathbb{R}$) et proposons plusieurs méthodes itératives qui ont pour but d'approcher la solution de

$$\begin{cases} \bar{x} \in [a, b], \\ f(\bar{x}) = 0. \end{cases} \quad (3.7)$$

Avant de procéder à la présentation de méthodes numériques pour approcher la solution de (3.7), nous rappelons le théorème des valeurs intermédiaires qui assure l'existence d'une solution dans l'intervalle $]a, b[$.

Théorème 3.3.1 (Théorème des valeurs intermédiaires) *Supposons que $f : [a, b] \mapsto \mathbb{R}$ est une fonction continue et k un réel entre $f(a)$ et $f(b)$. Alors il existe un réel $\xi \in]a, b[$ tel que $k = f(\xi)$. En particulier, si $f(a)f(b) \leq 0$, alors l'équation (3.7) admet au moins une solution.*

Commençons par présenter la méthode de dichotomie qui est certainement la plus naturelle mais pas forcément la plus efficace en termes de vitesse de convergence. Nous proposons ensuite l'algorithme de la sécante et poursuivons par la méthode de Newton-Raphson.

Dans cette partie, supposons que $f : [a, b] \mapsto \mathbb{R}$ est une fonction continue telle que $f(a)f(b) < 0$. Ainsi, puisque f est continue il existe forcément $\bar{x} \in]a, b[$ tel que $f(\bar{x}) = 0$.

³Ici cela revient à supposer que K est fermé, borné et convexe dans E .

3.3.1 Méthode de dichotomie

La méthode de dichotomie est la plus simple et sûrement la plus intuitive. Elle consiste à encadrer la solution \bar{x} par un intervalle de plus en plus petit. Nous posons $m = (a + b)/2$ et calculons $f(m)$. Puisque f est continue, testons le signe de la quantité $f(a)f(m)$. Ainsi, si $f(a)f(m) < 0$, ou autrement dit $f(a)$ et $f(m)$ sont de signes différents, cela signifie qu'il existe au moins un $\bar{x} \in [a, m]$ solution de $f(\bar{x}) = 0$. En revanche si $f(a)f(m) > 0$, ou autrement dit $f(a)$ et $f(m)$ sont de même signe, alors $f(m)f(b) < 0$ et par continuité de f , il existe $\bar{x} \in [m, b]$ tel que $f(\bar{x}) = 0$.

En itérant ce procédé, nous obtenons ainsi l'algorithme suivant : pour $\varepsilon > 0$ un petit paramètre,

Algorithme 2. Méthode de dichotomie

Poser $a^{(0)} = a$, $b^{(0)} = b$, $\varepsilon^{(0)} = b^{(0)} - a^{(0)}$ et $k = 0$.

Tant que $\varepsilon^{(k)} > \varepsilon$

- poser

$$m := \frac{a^{(k)} + b^{(k)}}{2},$$

si $f(a^{(k)})f(m) < 0$, alors

$$a^{(k+1)} = a^{(k)} \quad \text{et} \quad b^{(k+1)} = m,$$

sinon

$$a^{(k+1)} = m \quad \text{et} \quad b^{(k+1)} = b^{(k)},$$

- calculer $\varepsilon^{(k+1)} = b^{(k+1)} - a^{(k+1)}$ et $k \leftarrow k + 1$.

Fin de tant que

Observons bien qu'à chaque itération l'intervalle encadrant la solution \bar{x} telle que $f(\bar{x}) = 0$ est divisé par deux, soit par récurrence $(b^{(k)} - a^{(k)}) = (b^{(0)} - a^{(0)})/2^k$. Il en résulte que la méthode de dichotomie converge puisque $b^{(k)} - a^{(k)}$ tend vers zéro lorsque k tend vers l'infini. Nous pouvons donc choisir le temps d'arrêt N tel que

$$\frac{1}{2^N} (b^{(0)} - a^{(0)}) \leq \varepsilon,$$

où ε est la précision choisie.

Cet algorithme paraît très intéressant mais nous verrons qu'en fait cette méthode converge plutôt lentement. Néanmoins elle présente l'avantage d'être très simple et nécessite seulement que la fonction f soit continue.

3.3.2 Méthode de la sécante

Une autre méthode consiste simplement à remplacer f par une fonction affine approchée dans l'équation (3.7). En effet, prenons $a^{(0)}$ et $b^{(0)} \in [a, b]$ et plaçons-nous sur l'intervalle $[a^{(0)}, b^{(0)}]$, puis remplaçons f par un polynôme y de degré égal à un

$$y(x) := f(a^{(0)}) + (x - a^{(0)}) \frac{f(b^{(0)}) - f(a^{(0)})}{b^{(0)} - a^{(0)}},$$

où $y(x)$ est la fonction affine telle que $y(a^{(0)}) = f(a^{(0)})$ et $y(b^{(0)}) = f(b^{(0)})$. L'avantage est que pour une fonction affine y , l'équation $y(x) = 0$ peut être résolue exactement, pour tout $x \in [a^{(0)}, b^{(0)}]$. En effet, nous vérifions que $y(x^{(1)}) = 0$ avec

$$x^{(1)} = a^{(0)} - f(a^{(0)}) \frac{b^{(0)} - a^{(0)}}{f(b^{(0)}) - f(a^{(0)})}.$$

Géométriquement, ceci revient à remplacer la courbe $y = f(x)$ par la droite passant par les points $(a^{(0)}, f(a^{(0)}))$ et $(b^{(0)}, f(b^{(0)}))$ tandis que $x^{(1)}$ est simplement l'intersection de l'axe des abscisses avec cette droite. Pour trouver une meilleure approximation que $x^{(1)}$, il suffit de répéter le procédé soit sur l'intervalle $[a^{(0)}, x^{(1)}]$ ou $[x^{(1)}, b^{(0)}]$ selon le signe de $f(a^{(0)}) f(x^{(1)})$ comme pour la méthode de dichotomie.

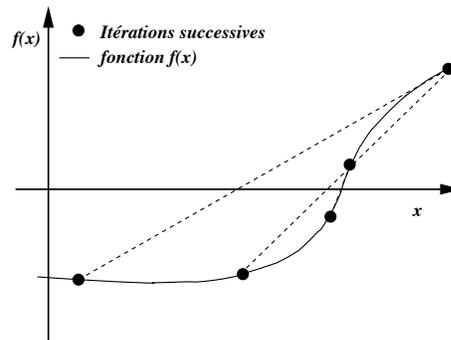


FIG. 3.2 – Illustration de la méthode de la sécante pour approcher la solution de $f(\bar{x}) = 0$.

Cette méthode peut se révéler plus efficace puisqu'à chaque étape nous réduisons la taille de l'intervalle bien plus vite que pour la méthode de dichotomie.

Finalement, pour une fonction $f : [a, b] \mapsto \mathbb{R}$ vérifiant $f(a) f(b) < 0$ et pour un petit paramètre $\varepsilon > 0$, l'algorithme est donné ci-dessous.

Algorithme 3. Méthode de la sécante

Poser $a^{(0)} = a$, $b^{(0)} = b$, $k = 0$ et $\varepsilon^{(0)} = b^{(0)} - a^{(0)}$.

Tant que $\varepsilon^{(k)} \geq \varepsilon$

$$\text{- calculer } x^{(k)} := a^{(k)} - f(a^{(k)}) \frac{a^{(k)} - b^{(k)}}{f(a^{(k)}) - f(b^{(k)})}$$

si $f(a^{(k)}) f(x^{(k)}) < 0$, alors

$$a^{(k+1)} = a^{(k)} \quad \text{et} \quad b^{(k+1)} = x^{(k)},$$

sinon

$$a^{(k+1)} = x^{(k)} \quad \text{et} \quad b^{(k+1)} = b^{(k)}.$$

- calculer le résidu $\varepsilon^{(k+1)} = b^{(k+1)} - a^{(k+1)}$ et $k \leftarrow k + 1$.

Fin de tant que.

Donnons alors un résultat de convergence pour la méthode de la sécante.

Théorème 3.3.2 (Convergence de la méthode de la sécante) Soient $f \in \mathcal{C}^0([a, b], \mathbb{R})$ vérifiant $f(a) f(b) < 0$ et $\bar{x} \in [a, b]$ une solution de (3.7). Alors en posant $a^{(0)} = a$ et $b^{(0)} = b$, la suite $(x^{(k)})_{k \in \mathbb{N}}$ donnée par la méthode de la sécante est bien définie et converge vers $\bar{x} \in [a, b]$.

3.3.3 Méthode de Newton-Raphson

La méthode de Newton-Raphson⁴ nécessite quelques hypothèses supplémentaires sur la fonction f , il faut notamment être en mesure de calculer sa dérivée.

Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction deux fois dérivable dont les dérivées sont continues, c'est-à-dire que $f \in \mathcal{C}^2([a, b], \mathbb{R})$. Nous cherchons à calculer numériquement les solutions de l'équation non linéaire $f(x) = 0$. Considérons $x_0 \in [a, b]$, une valeur approchée de la solution \bar{x} de (3.7). Posons alors $x^{(0)} = x_0$, puis en effectuant un développement de Taylor, nous obtenons

$$0 = f(\bar{x}) = f(x^{(0)}) + f'(x^{(0)}) (\bar{x} - x^{(0)}) + \frac{1}{2} f''(\eta) (\bar{x} - x^{(0)})^2,$$

⁴En référence à Isaac Newton et son contemporain Joseph Raphson. Cette méthode fut décrite de manière indépendante par les mathématiciens anglais Isaac Newton (1643-1727) et Joseph Raphson (1648-1715). Toutefois, Isaac Newton n'appliqua cet algorithme que pour la recherche de zéros d'équations polynomiales (voir Exercice 3.4.4), ce qui est assez simple. Notons qu'en 1690, Joseph Raphson en publia une description simplifiée, mais ce n'est qu'en 1740 que Thomas Simpson (1710-1761) décrivit cette méthode de calcul itératif pour approcher les solutions d'une équation non linéaire.

avec $\eta \in [a, b]$. En négligeant le terme d'ordre deux $|\bar{x} - x^{(0)}|^2$ qui sera petit dès que f'' est bornée et $x^{(0)}$ suffisamment proche de la solution \bar{x} , nous calculons une nouvelle approximation $x^{(1)}$ de \bar{x} donnée par

$$x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}.$$

Ainsi, en réitérant le procédé, nous obtenons un schéma itératif appelé méthode de Newton-Raphson

$$\begin{cases} x^{(0)} = x_0, \\ x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}, \quad k \geq 0. \end{cases} \quad (3.8)$$

Nous en déduisons l'algorithme suivant : pour $\varepsilon > 0$ un petit paramètre

Algorithme 4. Méthode de Newton-Raphson

Soit $x^{(0)} \in [a, b]$. Poser $k = 0$ et $\varepsilon^{(0)} = b - a$.

Tant que $\varepsilon^{(k)} \geq \varepsilon$

- calculer $x^{(k+1)} := x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}$

- calculer le résidu $\varepsilon^{(k+1)} = |x^{(k+1)} - x^{(k)}|$

et $k \leftarrow k + 1$.

Fin de tant que.

Pour cette méthode nous démontrons le théorème de convergence locale suivant.

Théorème 3.3.3 (Convergence locale de la méthode de Newton-Raphson) Soient $f \in \mathcal{C}^2([a, b], \mathbb{R})$ et $\bar{x} \in [a, b]$ une solution de (3.7) telle que $f'(\bar{x})$ soit non nul. Alors il existe $\delta > 0$ tel que pour tout $x_0 \in [\bar{x} - \delta, \bar{x} + \delta]$ la suite $(x^{(k)})_{k \in \mathbb{N}}$ donnée par (3.8) est bien définie et converge vers $\bar{x} \in [a, b]$. En outre, il existe une constante $C > 0$ telle que $|x^{(k+1)} - \bar{x}| \leq C |x^{(k)} - \bar{x}|^2$. Nous disons que la méthode de Newton-Raphson est au moins d'ordre deux.

Démonstration : Comme f est de classe $\mathcal{C}^2([a, b], \mathbb{R})$ et $f'(\bar{x})$ est différent de zéro, il existe $\delta > 0$, $L > 0$ et $M > 0$ tels que pour tout $x \in [\bar{x} - \delta, \bar{x} + \delta]$, les valeurs $f'(x)$ sont différentes de zéro et

$$\frac{1}{|f'(x)|} \leq \frac{1}{L}, \quad |f''(x)| \leq M.$$

En effet, en utilisant la continuité de f' au point \bar{x} , pour tout $\varepsilon > 0$, il existe $\delta > 0$, tel que pour tout $x \in [\bar{x} - \delta, \bar{x} + \delta]$, $|f'(x) - f'(\bar{x})| \leq \varepsilon$, alors en notant $\alpha = f'(\bar{x})$ que nous supposons par exemple strictement positif et en prenant $\varepsilon = \alpha/2$, nous obtenons $\alpha/2 \leq f'(x) \leq 3\alpha/2$. Donc, en prenant $L = \alpha/2$, nous avons pour tout $x \in [\bar{x} - \delta, \bar{x} + \delta]$, $|f'(x)| \geq L$. De plus, f étant de classe \mathcal{C}^2 sur l'intervalle $[\bar{x} - \delta, \bar{x} + \delta]$, il existe $M > 0$ tel que $|f''(x)| \leq M$.

Ainsi, en prenant au besoin une valeur de $\delta > 0$ plus petite encore, nous pouvons choisir $\delta > 0$ qui vérifie $M\delta/(2L) < 1$. En prenant $x^{(0)} \in]\bar{x} - \delta, \bar{x} + \delta[$ et en effectuant un développement de Taylor au voisinage du point $x^{(0)}$, nous obtenons

$$f(\bar{x}) = f(x^{(0)}) + f'(x^{(0)})(\bar{x} - x^{(0)}) + \frac{1}{2}f''(\eta^{(0)})(\bar{x} - x^{(0)})^2,$$

où $\eta^{(0)} \in [\bar{x} - \delta, \bar{x} + \delta]$. Puis, en écrivant

$$\begin{aligned} x^{(1)} &= x^{(0)} - [f'(x^{(0)})]^{-1} (f(x^{(0)}) - f(\bar{x})), \\ &= x^{(0)} + (\bar{x} - x^{(0)}) - \frac{1}{2} \frac{f''(\eta^{(0)})}{f'(x^{(0)})} (\bar{x} - x^{(0)})^2, \\ &= \bar{x} - \frac{1}{2} \frac{f''(\eta^{(0)})}{f'(x^{(0)})} (\bar{x} - x^{(0)})^2. \end{aligned}$$

Ainsi, en utilisant une majoration de $|f''|$ et une minoration de $|f'|$, nous avons finalement

$$|x^{(1)} - \bar{x}| \leq \frac{M}{2L} |x^{(0)} - \bar{x}|^2 \leq \frac{M}{2L} \delta^2 \leq \delta.$$

Nous procédons ensuite par récurrence. Supposons que $|x^{(k)} - \bar{x}| \leq \delta$, par un calcul identique au précédent, il vient

$$|x^{(k+1)} - \bar{x}| \leq \frac{M}{2L} |x^{(k)} - \bar{x}|^2 \leq \delta. \quad (3.9)$$

Ainsi, en posant

$$e^{(k)} = \frac{M}{2L} |x^{(k)} - \bar{x}|,$$

nous obtenons en utilisant (3.9)

$$e^{(k)} \leq [e^{(k-1)}]^2 \leq [e^{(k-2)}]^{2^2} \leq \dots \leq [e^{(0)}]^{2^k}.$$

Or, en ayant choisi au préalable $x^{(0)}$ tel que $e^{(0)} = |x^{(0)} - \bar{x}| \leq M\delta/2L < 1$, nous montrons que la méthode de Newton-Raphson est localement convergente. \square

Remarque 3.3.1 *Observons que cette méthode donne la convergence locale de la suite $(x^{(k)})_{k \in \mathbb{N}}$ vers la solution $\bar{x} \in [a, b]$. Il faut donc partir d'un point $x^{(0)}$ suffisamment proche de \bar{x} , cela nécessite la connaissance a priori de la solution \bar{x} , ce qui n'est pas toujours possible !*

Nous venons de voir que sous certaines hypothèses, la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers la solution $\bar{x} \in [a, b]$. Néanmoins dans la pratique un critère d'arrêt est nécessaire pour stopper les itérations. Par exemple, le critère d'arrêt peut être $|x^{(k+1)} - x^{(k)}| < \varepsilon$, où $\varepsilon > 0$ est un petit paramètre fixé. Toutefois cela n'assure pas à chaque fois la convergence vers la solution \bar{x} . En effet, considérons par exemple la suite $x^{(k)} = \sum_{l=1}^k 1/l$ nous avons

$$\lim_{k \rightarrow +\infty} |x^{(k+1)} - x^{(k)}| = 0,$$

mais $x^{(k)}$ tend vers l'infini. Remarquons qu'un autre choix possible est le critère $|f(x^{(k)})| < \varepsilon$.

Exemple 3.3.1 Prenons $f(x) = x e^{-x^2}$, nous avons alors $f'(x) = (1 - 2x^2) e^{-x^2}$. L'algorithme de la méthode de Newton-Raphson s'écrit $x^{(k+1)} = x^{(k)} - x^{(k)}/(1 - 2|x^{(k)}|^2)$.

D'une part, en prenant $x^{(0)} = 0,3$, nous obtenons :

k	$x^{(k)}$	$f(x^{(k)})$
0	0,3	0,274
1	$-6,6 \cdot 10^{-02}$	$-6,6 \cdot 10^{-02}$
2	$+5,8 \cdot 10^{-04}$	$+5,8 \cdot 10^{-04}$
3	$-3,8 \cdot 10^{-10}$	$-3,8 \cdot 10^{-10}$
4	$+1,1 \cdot 10^{-28}$	$+1,1 \cdot 10^{-28}$
5	$-2,8 \cdot 10^{-84}$	×

D'autre part, en prenant $x^{(0)} = 0,5$, la méthode ne converge pas mais $(x^{(k)})_k$ prend seulement les valeurs $\pm 0,5$.

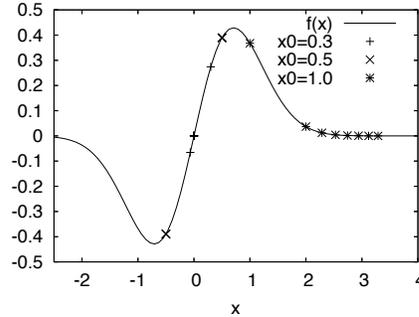


FIG. 3.3 – La fonction $f(x)$ et des valeurs $(x^{(k)})_{k \in \mathbb{N}}$ pour des données initiales différentes : $x^{(0)} = 0,3$ (+), $x^{(0)} = 0,5$ (x) et $x^{(0)} = 1$ (*).

Enfin pour $x^{(0)} = x_0 = 1$, la suite $(x^{(k)})_k$ converge vers $+\infty$ lorsque k tend vers l'infini.

3.3.4 Combinaison de méthodes

La méthode de Newton-Raphson est incontestablement la meilleure parmi celles présentées dans ce chapitre mais à la seule condition de choisir la donnée initiale x_0 assez proche de la solution recherchée, où δ assez proche de ζ dépend de la fonction f . Plus précisément, lorsque nous connaissons bien les dérivées premières et secondes de f au voisinage de la solution, nous pouvons combiner deux méthodes de façon pragmatique.

L'algorithme est simplement donné, pour un petit paramètre $\varepsilon > 0$, par

Algorithme 5. Algorithme de combinaison de méthodes

Poser $a^{(0)} = a$ et $b^{(0)} = b$ tels que $f(a^{(0)}) f(b^{(0)}) < 0$

et $\varepsilon^{(0)} = b^{(0)} - a^{(0)}$, $k = 0$.

Tant que $\varepsilon^{(k)} > \varepsilon$

- calculer $x^{(k+1)}$ à l'aide de la méthode de Newton-Raphson :

- si $x^{(k+1)} \notin [a^{(k)}, b^{(k)}]$, le rejeter et effectuer

une itération de dichotomie pour trouver $[a^{(k+1)}, b^{(k+1)}]$.

- si $x^{(k+1)} \in [a^{(k)}, b^{(k)}]$, alors :

si $f(x^{(k+1)}) f(a^{(k)}) < 0$, poser

$$a^{(k+1)} = a^{(k)} \text{ et } b^{(k+1)} = x^{(k+1)}.$$

si $f(x^{(k+1)}) f(b^{(k)}) < 0$, poser

$$a^{(k+1)} = x^{(k+1)} \text{ et } b^{(k+1)} = b^{(k)}.$$

- calculer $\varepsilon^{(k+1)} = b^{(k+1)} - a^{(k+1)}$ et $k \leftarrow k + 1$.

Fin de tant que

3.4 Méthode de Newton-Raphson dans \mathbb{R}^n

Avant de procéder à la présentation de la méthode de Newton-Raphson pour des fonctions de \mathbb{R}^n à valeurs dans \mathbb{R}^n , nous rappelons quelques notions de calcul différentiel.

3.4.1 Quelques rappels de calcul différentiel

Considérons deux \mathbb{R} -espaces de Banach E et F . Notons $\mathcal{L}(E, F)$ l'espace des applications linéaires continues de E dans F , muni de la norme

$$\|l\| = \sup_{h \in E} \frac{\|lh\|_F}{\|h\|_E},$$

laquelle correspond à la norme matricielle lorsque $E = F = \mathbb{R}^n$. Observons que $\mathcal{L}(E, F)$ est également un espace de Banach.

Supposons que U est un ouvert (non vide !) de E et considérons une fonction $f : U \mapsto F$.

Définition 3.4.1 *Nous disons que la fonction f est différentiable en un point $x \in U$ si elle est continue au point x et s'il existe $l(x) \in \mathcal{L}(E, F)$ tel que*

$$\lim_{h \rightarrow 0_E} \frac{\|f(x+h) - f(x) - l(x)h\|_F}{\|h\|_E} = 0. \quad (3.10)$$

L'application $l(x)$ dépend du point x , nous la noterons désormais $l(x) = \nabla f(x)$, de sorte que (3.10) se réécrit

$$\lim_{h \rightarrow 0_E} \frac{\|f(x+h) - f(x) - \nabla f(x)h\|_F}{\|h\|_E} = 0.$$

L'application $\nabla f(x)$ est appelée différentielle de f au point x .

Définition 3.4.2 *Nous disons que la fonction f est différentiable sur U si elle est différentiable en tout point $x \in U$. Dans ce cas, nous appelons différentielle de f la fonction*

$$\begin{aligned} \nabla f : \quad & U \mapsto \mathcal{L}(E, F) \\ & x \mapsto \nabla f(x). \end{aligned}$$

Si de plus ∇f est continue par rapport à x , nous disons que f est continûment différentiable, ou de façon équivalente que f est de classe $\mathcal{C}^1(E, F)$.

Cas particulier $E = \mathbb{R}^p$ et $F = \mathbb{R}^q$

Pour tout $x \in U \subset E$, $x = (x_1, \dots, x_p)^T$ et pour tout $i \in \{1, \dots, p\}$, l'ensemble

$$V_i(x) := \{t \in \mathbb{R}, (x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_p) \in U\}$$

est un voisinage ouvert de x_i .

Supposons $f : U \subset E \mapsto F$ différentiable. Alors l'application partielle $g_i : V_i(x) \mapsto F$, donnée par $t \mapsto f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_p) = f(x + (t - x_i)e_i)$, est dérivable en x_i et $g'_i(x_i) = \nabla f(x) e_i$, c'est en fait la dérivée partielle de f dans la direction e_i au point x . Dans la suite, nous notons cette dérivée partielle $\frac{\partial f}{\partial x_i}(x)$ et

$$\begin{aligned} \frac{\partial f}{\partial x_i} : \quad & U \mapsto F \\ & x \mapsto \frac{\partial f}{\partial x_i}(x) \end{aligned}$$

pour $i \in \{1, \dots, p\}$. Par linéarité de $\nabla f(x)$, nous vérifions que pour tout $h \in \mathbb{R}^p$, $h = (h_1, \dots, h_p)^T$,

$$\nabla f(x)h = \sum_{i=1}^p h_i \frac{\partial f}{\partial x_i}(x).$$

Pour tout $i \in \{1, \dots, p\}$, l'application $h \in E \mapsto h_i \in \mathbb{R}$ est une forme linéaire continue, c'est-à-dire un élément de $\mathcal{L}(\mathbb{R}^p, \mathbb{R})$. Nous la notons dx_i . Ainsi, la différentielle de f au point x s'écrit

$$\nabla f(x) = \sum_{i=1}^p \frac{\partial f}{\partial x_i}(x) dx_i.$$

Notons bien que l'existence de dérivées partielles n'est pas suffisante en général pour qu'une fonction soit différentiable. En revanche,

Théorème 3.4.1 *Une application $f : U \subset E \mapsto F$ est continûment différentiable si et seulement si ses p dérivées partielles existent et sont continues sur U .*

Si une fonction $f : U \subset E \mapsto F$, de composantes (f_1, \dots, f_q) , est différentiable au point x , nous définissons sa matrice jacobienne au point x comme la matrice de l'application linéaire $\nabla f(x)$ dans les bases canoniques de E et F . Elle est donnée par

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_p}(x) \\ \vdots & & \vdots \\ \frac{\partial f_q}{\partial x_1}(x) & \dots & \frac{\partial f_q}{\partial x_p}(x) \end{pmatrix} \in \mathcal{M}_{q,p}(\mathbb{R}).$$

Autrement dit, le coefficient de la matrice jacobienne de f d'indice $i \in \{1, \dots, q\}$ en ligne et $j \in \{1, \dots, p\}$ en colonne est $(\nabla f(x))_{i,j} = \frac{\partial f_i}{\partial x_j}(x)$.

En particulier, si $q = 1$, $\nabla f(x)$ est une matrice ligne. De façon générale, les lignes des $\nabla f(x)$ sont les $\nabla f_i(x)$.

Cas général

Supposons maintenant que E et F sont des espaces de Banach. Définissons alors les fonctions lipschitziennes de la façon suivante :

Théorème 3.4.2 (Fonction lipschitzienne) *Soit $f : U \subset E \mapsto F$ une fonction différentiable sur un ouvert convexe U . Supposons que sa différentielle est bornée, c'est-à-dire il existe $k > 0$ tel que $\|\nabla f(u)\| \leq k$, pour tout $u \in U$. Alors la fonction f est lipschitzienne, c'est-à-dire que $\|f(x) - f(y)\|_F \leq k\|x - y\|_E$, pour tout $(x, y) \in U \times U$*

Remarque 3.4.1 *La démonstration donne en fait l'inégalité plus fine*

$$\|f(y) - f(x)\|_F \leq \sup_{t \in [0,1]} \|\nabla f(x + t(y - x))\| \|y - x\|_E.$$

Théorèmes d'inversion locale et des fonctions implicites Soient U et V des ouverts (non vides) des espaces de Banach E et F respectivement.

Définition 3.4.3 *Une application $f : U \mapsto V$ est un \mathcal{C}^1 -difféomorphisme de U sur V si et seulement si*

- la fonction f est une bijection,

- la fonction f est de classe $\mathcal{C}^1(U, V)$, c'est-à-dire continûment différentiable sur U ,
- la fonction inverse f^{-1} est de classe $\mathcal{C}^1(V, U)$.

Le théorème d'inversion locale est le suivant :

Théorème 3.4.3 (Théorème d'inversion locale) *Si $f : U \mapsto V$ est une fonction de classe $\mathcal{C}^1(U, V)$, si $a \in U$ est tel que $\nabla f(a)$ soit un isomorphisme de E sur F , alors il existe un voisinage ouvert U_a de a dans U et un voisinage ouvert V_b de $b = f(a)$ dans V tel que la restriction de f à U_a soit un \mathcal{C}^1 -difféomorphisme de U_a sur V_b .*

Parmi les conséquences fondamentales du théorème d'inversion locale, nous trouvons un résultat tout aussi important, connu sous le nom de théorème des fonctions implicites. Il concerne le problème qui nous occupe dans ce chapitre, c'est-à-dire la résolution d'équations non linéaires.

Considérons trois espaces de Banach E, F et G et $f : E \times F \mapsto G$ et recherchons les solutions $(x, y) \in E \times F$ de l'équation $f(x, y) = 0_G$. Sous des hypothèses que nous allons préciser, nous pouvons en tirer y comme fonction de x : nous disons alors que $f(x, y) = 0$ définit implicitement y , ou encore y comme fonction implicite de x . Plus précisément, nous avons :

Théorème 3.4.4 (Théorème des fonctions implicites) *Soit U un ouvert de $E \times F$ et*

$$f : \begin{array}{l} U \quad \mapsto \quad G \\ (x, y) \mapsto f(x, y), \end{array}$$

une fonction de classe $\mathcal{C}^1(U, G)$ et notons $\nabla f = (\nabla_x f, \nabla_y f)$ le gradient de la fonction f par rapport à $x \in E$ et $y \in F$. Supposons qu'il existe $(a, b) \in U$ tel que $f(a, b) = 0_G$ et $\nabla_y f(a, b) \in \mathcal{L}(F, G)$ est un isomorphisme de F sur G . Alors il existe un voisinage ouvert $U(a, b)$ de (a, b) dans U , un voisinage ouvert W_a de a dans E et une fonction $\varphi \in \mathcal{C}^1(W_a, F)$ tels que

$$(x, y) \in U(a, b), \quad \text{et} \quad f(x, y) = 0_G \Leftrightarrow y = \varphi(x).$$

Dérivées d'ordre supérieur et développements de Taylor

Commençons par définir une fonction deux fois différentiable puis enchaînons sur les développements de Taylor à l'ordre supérieur.

Définition 3.4.4 *Une fonction f définie sur un ouvert (non vide) U d'un \mathbb{R} -espace de Banach E et à valeurs dans un \mathbb{R} -espace de Banach F est dite deux fois différentiable en $x \in U$ si elle est différentiable dans un voisinage ouvert U_x de x et si sa différentielle $\nabla f : U_x \mapsto \mathcal{L}(E, F)$ est différentiable en x . Nous disons que f est deux fois différentiable dans U si elle est deux fois différentiable en tout point de U .*

Par définition, la différentielle de ∇f en x , $\nabla^2 f := \nabla(\nabla f)(x)$ est une application linéaire continue de E dans $\mathcal{L}(E, F)$. De même, définissons de manière récursive les espaces de fonctions p -fois continûment différentiables, notés $\mathcal{C}^p(E, F)$ et la différentielle d'ordre p d'une fonction f , notée $\nabla^p f$ puis $\nabla^{p+1} f = \nabla(\nabla^p f)$.

Théorème 3.4.5 (Développement de Taylor avec reste intégral) Soient E, F deux espaces de Banach, U un ouvert de E et $f : U \mapsto F$ une fonction de classe \mathcal{C}^{n+1} . Alors pour tout $(x, h) \in U \times E$ tel que le segment $[x, x + h]$ soit inclus dans U ,

$$f(x + h) = f(x) + \sum_{p=1}^n \frac{1}{p!} \nabla^p f(x) h^{[p]} + \int_0^1 \frac{(1-t)^n}{n!} \nabla^{n+1} f(x + th) h^{[n+1]} dt.$$

$$\text{où } h^{[p]} = \underbrace{(h, \dots, h)}_{p \text{ fois}}.$$

Présentons également une autre version de la formule de Taylor, valable sous des hypothèses encore moins fortes, et qui pour cette raison donne un résultat local seulement.

Théorème 3.4.6 Soient E, F deux espaces de Banach, U un ouvert de E et $f : U \mapsto F$ une fonction n fois différentiable en $x \in U$. Alors

$$\|f(x + h) - f(x) - \sum_{p=1}^n \frac{1}{p!} \nabla^p f(x) h^{[p]}\| = o(\|h\|^n),$$

où $o(\|h\|^n)$ signifie qu'il existe une fonction $\epsilon(h)$ telle que $\epsilon(h)/\|h\|^n \rightarrow 0$ lorsque h tend vers 0.

3.4.2 Méthode de Newton-Raphson

Soient U un ouvert de \mathbb{R}^n et $f \in \mathcal{C}^0(U, \mathbb{R}^n)$. Supposons dans cette partie qu'il existe $\bar{x} \in U$ solution de $f(\bar{x}) = 0$. Nous souhaitons mettre au point une méthode numérique pour approcher cette solution. Pour cela, nous procédons de manière analogue à ce que nous avons fait pour des fonctions à valeurs réelles.

Tout d'abord, nous choisissons une première approximation $x_0 \in U$ de \bar{x} et remplaçons ensuite la fonction f par son développement de Taylor à l'ordre un au point x_0 , puis en itérant ce procédé nous obtenons une suite $(x^{(k)})_{k \in \mathbb{N}}$ donnée par

$$\begin{cases} x^{(0)} = x_0 \in U, \\ \\ f(x^{(k)}) + \nabla f(x^{(k)}) (x^{(k+1)} - x^{(k)}) = 0_{\mathbb{R}^n}, \quad k \geq 0. \end{cases} \quad (3.11)$$

Pour chaque $k \in \mathbb{N}$, nous devons donc résoudre le problème (3.11). Il nous faut donc :

- calculer la matrice $A = \nabla f(x^{(k)})$,
- s'assurer que cette matrice $A \in \mathcal{M}_{n,n}(\mathbb{R})$ est bien inversible,
- résoudre le système $A x^{(k+1)} = -f(x^{(k)}) + A x^{(k)} \in \mathbb{R}^n$,
- s'assurer que l'itéré suivant $x^{(k+1)}$ appartient bien à U pour pouvoir continuer l'algorithme.

Lorsque $n = 1$, nous avons facilement montré que la méthode de Newton-Raphson est convergente et d'ordre deux. Nous allons voir que ce résultat est toujours vrai en dimension $n \geq 1$.

Théorème 3.4.7 (Convergence locale de la méthode de Newton-Raphson dans \mathbb{R}^n) Considérons $f : U \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ une fonction deux fois différentiable et dont la différentielle d'ordre deux est continue, c'est-à-dire que f est de classe $\mathcal{C}^2(U, \mathbb{R}^n)$.

Soit $\bar{x} \in U$ tel que $f(\bar{x}) = 0$ et la matrice $\nabla f(\bar{x})$ est inversible. Alors :

- il existe $\delta > 0$ tel que pour tout $x_0 \in \mathcal{B}(\bar{x}, \delta)$, la suite $(x^{(k)})_{k \in \mathbb{N}}$ donnée par (3.11) est bien définie et $x^{(k)} \in \mathcal{B}(\bar{x}, \delta)$, pour tout $k \in \mathbb{N}$;
- la suite $(x^{(k)})_{k \in \mathbb{N}}$ donnée par (3.11) converge vers la solution $\bar{x} \in U$;
- il existe une constante $C > 0$ telle que $\|x^{(k+1)} - \bar{x}\| \leq C \|x^{(k)} - \bar{x}\|^2$.

Ici $\mathcal{B}(\bar{x}, \delta)$ représente la boule de \mathbb{R}^n de centre \bar{x} et de rayon δ .

Avant d'aborder la preuve de ce théorème, présentons trois lemmes intermédiaires : le premier est simplement un résultat technique sur les matrices, le deuxième fournit un résultat de stabilité de la fonction f au voisinage du point $\bar{x} \in U$ tel que $f(\bar{x}) = 0$ et enfin le troisième lemme assure l'existence d'une solution à chaque étape de l'algorithme de Newton-Raphson.

D'abord, proposons un rappel sur les matrices et les normes matricielles.

Lemme 3.4.1 Soit $B \in \mathcal{M}_{n,n}(\mathbb{K})$ une matrice dans un corps \mathbb{K} telle que $\|B\| < 1$, pour une norme matricielle subordonnée $\|\cdot\|$.

Alors la matrice $(I_n - B)$ est inversible et

$$(I_n - B)^{-1} = \sum_{k=0}^{+\infty} B^k, \quad \|(I_n - B)^{-1}\| \leq \frac{1}{1 - \|B\|}.$$

Démonstration : Remarquons tout d'abord que la condition $\|B\| < 1$ assure la convergence de la série puisque

$$\left\| \sum_{k=0}^{+\infty} B^k \right\| \leq \sum_{k=0}^{+\infty} \|B\|^k < +\infty.$$

En outre,

$$(I_n - B) \sum_{k=0}^p B^k = \sum_{k=0}^p (I_n - B) B^k = I_n - B^{p+1}$$

et puisque $\|B\| < 1$, nous obtenons en passant à la limite $p \rightarrow +\infty$

$$(I_n - B) \sum_{k=0}^{+\infty} B^k = I_n.$$

Donc $(I_n - B)^{-1} = \sum_{k=0}^{+\infty} B^k$. Enfin,

$$\|(I_n - B)^{-1}\| = \left\| \sum_{k=0}^{+\infty} B^k \right\| \leq \sum_{k=0}^{+\infty} \|B\|^k \leq \frac{1}{1 - \|B\|}.$$

□

Ensuite, mettons en évidence un résultat de régularité de la fonction f au voisinage du point $\bar{x} \in U$ vérifiant $f(\bar{x}) = 0$.

Lemme 3.4.2 Soient $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction de classe $\mathcal{C}^2(U, \mathbb{R}^n)$ et $\bar{x} \in U$ tel que $f(\bar{x}) = 0$ et $\nabla f(\bar{x})$ est inversible. Alors il existe $L > 0$, $M > 0$ et un paramètre $\delta > 0$ vérifiant $M\delta/L < 1$, tels que pour tout $x \in \mathcal{B}(\bar{x}, \delta)$, $\|[\nabla f(x)]^{-1}\| \leq 1/L$ et de plus $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$.

Démonstration : Puisque $f \in \mathcal{C}^2(U, \mathbb{R}^n)$, pour tout $x \in U$ le terme $\|\nabla^2 f(x)\|$ est borné sur tout sous-ensemble borné de U et d'après le Théorème 3.4.2, pour un rayon $r > 0$ fixé, il existe $M(r) > 0$ tel que pour tout $x, y \in \mathcal{B}(\bar{x}, r) \subset U$,

$$\|\nabla f(x) - \nabla f(y)\| \leq M(r) \|x - y\|.$$

Remarquons bien que r est fixé et quelconque ; il reste donc à montrer qu'il existe $\delta > 0$ et $L > 0$ tels que pour tout $x \in \mathcal{B}(\bar{x}, \delta)$, la matrice $\nabla f(x)$ est inversible et vérifie

$$\|\nabla f(x)^{-1}\| \leq \frac{1}{L}.$$

Pour cela, nous écrivons

$$\begin{aligned} \nabla f(x) &= \nabla f(\bar{x}) - \nabla f(\bar{x}) + \nabla f(x), \\ &= \nabla f(\bar{x}) \left[I_n - [\nabla f(\bar{x})]^{-1} (\nabla f(\bar{x}) - \nabla f(x)) \right] \end{aligned}$$

et posons $B = [\nabla_x f(\bar{x})]^{-1} (\nabla f(\bar{x}) - \nabla f(x))$, ce qui s'écrit alors $\nabla f(x) = \nabla f(\bar{x}) [I_n - B]$. Ainsi, puisque $\nabla f(\bar{x})$ est inversible, la matrice $\nabla f(x)$ est inversible dès que $[I_n - B]$ est elle-même inversible. En vue d'appliquer le Lemme 3.4.1, il nous suffit de montrer qu'il existe $\delta_0 > 0$ tel que pour tout $x \in \mathcal{B}(\bar{x}, \delta_0)$, nous avons $\|B\| < 1$. Or d'après ce qui précède nous avons

$$\|B\| \leq \|[\nabla f(\bar{x})]^{-1}\| \|\nabla f(x) - \nabla f(\bar{x})\| \leq \|[\nabla f(\bar{x})]^{-1}\| M(r) \|x - \bar{x}\|$$

et en prenant $\delta_0 > 0$ suffisamment petit tel que $M(r) \|[\nabla_x f(\bar{x})]^{-1}\| \delta_0 < 1$, nous avons pour tout $x \in \mathcal{B}(\bar{x}, \delta_0)$, $\|B\| < 1$, ce qui signifie que pour tout $x \in \mathcal{B}(\bar{x}, \delta_0)$, la matrice $\nabla f(x)$ est inversible et

$$\begin{aligned} \|[\nabla f(x)]^{-1}\| &\leq \|[\nabla f(\bar{x})]^{-1}\| \|(I_n - B)^{-1}\|, \\ &\leq \|[\nabla f(\bar{x})]^{-1}\| \frac{1}{1 - \|B\|} =: \frac{1}{L}. \end{aligned}$$

Nous choisissons finalement $M := M(r)$ et $\delta \in]0, \delta_0[$ tel que $M(r) \delta / L < 1$, ce qui conclut la démonstration. \square

Enfin, le troisième lemme assure l'existence de la suite $(x^{(k)})_{k \in \mathbb{N}}$ obtenue à l'aide de l'algorithme de Newton-Raphson.

Lemme 3.4.3 Soient ω un ouvert de \mathbb{R}^n et $f : \omega \mapsto \mathbb{R}^n$ une fonction telle que $f \in \mathcal{C}^1(\omega, \mathbb{R}^n)$, $\nabla f(x)$ est inversible pour tout $x \in \omega$ et il existe $L > 0$ et $M > 0$ vérifiant

$$\|[\nabla f(x)]^{-1}\| \leq \frac{1}{L}, \quad \|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\| \quad (3.12)$$

et pour $k \geq 1$, $x^{(k-1)} \in \omega$.

Alors le terme $x^{(k)}$ donné par la méthode de Newton-Raphson est bien défini et tel que

$$\|x^{(k)} - x^{(k-1)}\| \leq \frac{1}{L} \|f(x^{(k-1)})\| \quad (3.13)$$

et

$$\|f(x^{(k)})\| \leq M \|x^{(k)} - x^{(k-1)}\|^2. \quad (3.14)$$

Démonstration : D'une part, puisque $x^{(k-1)} \in \omega$, la matrice jacobienne $\nabla f(x^{(k-1)})$ est inversible, nous pouvons donc construire $x^{(k)} = x^{(k-1)} - [\nabla f(x^{(k-1)})]^{-1} f(x^{(k-1)})$, d'où nous déduisons, grâce à l'hypothèse (3.12), que

$$\|x^{(k)} - x^{(k-1)}\| = \|[\nabla f(x^{(k-1)})]^{-1} f(x^{(k-1)})\| \leq C \|f(x^{(k-1)})\|.$$

D'autre part, à l'aide de la formule de Taylor-Young avec reste intégral, nous avons aussi

$$f(x^{(k)}) = f(x^{(k-1)}) + \int_0^1 \nabla f(x^{(k-1)} + t(x^{(k)} - x^{(k-1)})) (x^{(k)} - x^{(k-1)}) dt.$$

Puis, en utilisant la méthode de Newton-Raphson à l'étape $k - 1$, $f(x^{(k-1)}) + \nabla f(x^{(k-1)}) (x^{(k)} - x^{(k-1)}) = 0$, nous avons

$$f(x^{(k)}) = \int_0^1 \left[\nabla f(x^{(k-1)} + t(x^{(k)} - x^{(k-1)})) - \nabla f(x^{(k-1)}) \right] (x^{(k)} - x^{(k-1)}) dt.$$

D'où,

$$\begin{aligned} \|f(x^{(k)})\| &\leq \sup_{t \in]0,1[} \left(\|\nabla f(x^{(k-1)} + t(x^{(k)} - x^{(k-1)})) - \nabla f(x^{(k-1)})\| \right) \|x^{(k)} - x^{(k-1)}\|, \\ &\leq \sup_{t \in]0,1[} \left(M t \|x^{(k)} - x^{(k-1)}\| \right) \|x^{(k)} - x^{(k-1)}\|, \end{aligned}$$

c'est-à-dire $\|f(x^{(k)})\| \leq M \|x^{(k)} - x^{(k-1)}\|^2$. \square

Nous sommes dorénavant en mesure de démontrer le Théorème 3.4.7.

Démonstration du Théorème 3.4.7. D'abord, en appliquant le Lemme 3.4.2, puisque la matrice $\nabla f(\bar{x})$ est inversible et $f \in \mathcal{C}^2(\bar{U}, \mathbb{R}^n)$, il existe $L > 0$, M et $\delta > 0$ vérifiant $M\delta/L\delta < 1$, tels que pour tout $x \in \mathcal{B}(\bar{x}, \delta)$, la matrice $\nabla f(x)$ est inversible et

$$\|[\nabla f(x)]^{-1}\| \leq \frac{1}{L}.$$

De plus, pour tout $x, y \in \mathcal{B}(\bar{x}, \delta)$, $\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|$.

Dans un premier temps, nous souhaitons démontrer que chaque terme $x^{(k)}$ est bien défini pour tout $k \in \mathbb{N}$ et ensuite que la suite $(f(x^{(k)}))_{k \geq 0}$ tend vers zéro lorsque k tend vers l'infini. Pour cela, nous raisonnons par récurrence.

Soit $x_0 \in \mathcal{B}(\bar{x}, \delta)$. Montrons que si le terme $x^{(k)} \in \mathcal{B}(\bar{x}, \delta)$ pour $k \geq 0$, alors nous avons aussi $x^{(k+1)} \in \mathcal{B}(\bar{x}, \delta)$. Tout d'abord en appliquant le Lemme 3.4.3, la valeur $x^{(k+1)}$ est bien définie. Ensuite à l'aide d'un développement de Taylor-Young à l'ordre deux de la fonction f et en rappelant que $f(\bar{x}) = 0$, il vient

$$\|f(x^{(k)}) + \nabla f(x^{(k)}) (\bar{x} - x^{(k)})\| \leq M \|\bar{x} - x^{(k)}\|^2.$$

Aussi, en utilisant la définition de $x^{(k+1)}$, nous avons $\|\nabla f(x^{(k)}) (x^{(k+1)} - \bar{x})\| \leq M \|\bar{x} - x^{(k)}\|^2$. Puis,

$$\begin{aligned} \|x^{(k+1)} - \bar{x}\| &= \|\nabla f(x^{(k)})^{-1} \nabla f(x^{(k)}) (x^{(k+1)} - \bar{x})\| \\ &\leq \|\nabla f(x^{(k)})^{-1}\| \|\nabla f(x^{(k)}) (x^{(k+1)} - \bar{x})\| \\ &\leq \frac{M}{L} \|x^{(k)} - \bar{x}\|^2 \end{aligned} \tag{3.15}$$

et donc par définition de δ , nous avons $\|x^{(k+1)} - \bar{x}\| \leq M \delta^2/L \leq \delta$, ce qui montre bien que $x^{(k+1)} \in \mathcal{B}(\bar{x}, \delta)$ et donc le premier point du Théorème 3.4.7.

Notons que nous avons par la même occasion démontré l'existence d'une constante $C = M/L > 0$ telle que $\|x^{(k+1)} - \bar{x}\| \leq C \|x^{(k)} - \bar{x}\|^2$, ce qui prouve la dernière assertion du Théorème 3.4.7.

Intéressons-nous alors au deuxième point, c'est-à-dire à la convergence de la suite $(x^{(k)})_{k \geq 0}$. Pour cela posons $e^{(k)} := M \|x^{(k)} - \bar{x}\|/L$, il vient alors en multipliant (3.15) par M/L ,

$$e^{(k+1)} \leq [e^{(k)}]^2 \leq [e^{(k-1)}]^{2^2} \leq \dots \leq [e^{(0)}]^{2^{k+1}}.$$

Or, en ayant choisi au préalable $x^{(0)}$ tel que $e^{(0)} = \|x^{(0)} - \bar{x}\| \leq M \delta/L < 1$, nous montrons que la méthode de Newton-Raphson est bien localement convergente. \square

Ce théorème montre la convergence locale de la méthode de Newton-Raphson, c'est-à-dire que si x_0 est suffisamment proche d'une solution $f(\bar{x}) = 0$, alors la suite $(x^{(k)})_{k \in \mathbb{N}}$ converge vers \bar{x} . Concernant la convergence globale, nous en savons très peu et nous ne pouvons analyser que quelques cas de fonctions simples.

Un exemple classique de convergence globale est $f(\bar{z}) = 0$, où $\bar{z} \in \mathbb{C}$, avec

$$f(z) = z^3 - 1, \quad z \in \mathbb{C}$$

ou encore

$$f(x, y) = \begin{pmatrix} x^3 - 3xy^2 - 1 \\ 3x^2y - y^3 \end{pmatrix},$$

avec $z = x + iy$ et pour lequel l'itération devient

$$\begin{aligned} z^{(k+1)} &= z^{(k)} - \frac{(z^{(k)})^3 - 1}{3(z^{(k)})^2} \\ &= \frac{1}{3} \left(2z^{(k)} + \frac{1}{(z^{(k)})^2} \right). \end{aligned}$$

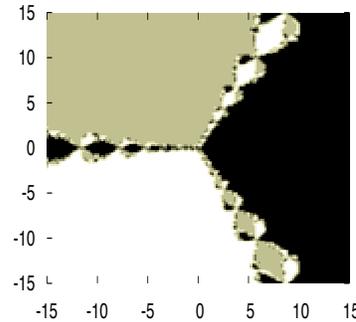


FIG. 3.4 – Données initiales convergent vers la solution $z = 1$ (noir), $z = -(1 - \sqrt{3})/2$ (gris) et $z = -(1 + \sqrt{3})/2$ (blanc).

Nous posons alors

$$A(a) = \left\{ z_0 \in \mathbb{C}, \quad (z^{(k)})_k \text{ converge vers } a \right\},$$

avec $a = 1, (-1 \pm i\sqrt{3})/2$ racines de $f(z) = 0$. Dans ce cas, nous vérifions à l'aide d'un ordinateur que pour toutes valeurs de z_0 , l'algorithme de Newton-Raphson converge vers une solution (voir pour cela la Figure 3.4).

Dans la pratique, il arrive souvent que la forme analytique de la matrice $\nabla f(x)$ est inconnue. Dans ce cas nous approchons les éléments $\partial f_i / \partial x_j$ de la matrice jacobienne par

$$\frac{\partial f_i}{\partial x_j}(x_1, \dots, x_n) \simeq \frac{f_i(x_1, \dots, x_j + \delta, \dots, x_n) - f_i(x_1, \dots, x_j, \dots, x_n)}{\delta},$$

en choisissant le paramètre δ de l'ordre de l'erreur d'arrondi.

Notons que beaucoup de méthodes dites quasi Newton pour la résolution numérique de systèmes linéaires consistent à trouver de nouveaux algorithmes qui ne nécessitent plus le calcul exact de la jacobienne. Nous renvoyons le lecteur à l'ouvrage de J. E. Dennis et R. B. Schnabel [12], qui dresse l'état de l'art des méthodes de Newton-Raphson et quasi-Newton tant du point de vue théorique que pratique.