



Support de cours

« Méthode Numérique » (2^{eme} année ST LMD)

Programme détaillé du module

Chapitre 01 : Résolution numériques des équations non linéaires a une variable

1.1 Localisation des racines d'une équation $f(x)=0$.

1.2 Méthode de Bissection où de Dichotomie :

1.3 Méthode des approximations successives où du point fixe

1.4 Méthode de Newton-Raphson :

Chapitre 2 : Interpolation polynômiale

2.1 Polynôme d'interpolation de Lagrange.

2.2 Polynôme d'interpolation de Newton.

2.3 Erreur d'interpolation

Chapitre 3 : Intégration Numérique

3.1 Formule du trapèze

3.1.1 Formule du trapèze généralisées.

3.1.2 Erreur d'intégration

3.2 Formule du Simpson.

3.3 Formules de quadratures :

Chapitre 4 : Résolution directe des systèmes d'équations linéaires

4.1 Méthode d'élimination de GAUSS

4.2 Algorithme de Thomas

4.3 Méthode de Crout-Dolittle ou LU

4.4 Méthode de Choleski

Chapitre 5 : Résolution Numérique des équations différentielles ordinaires : Problème de Cauchy

5.1 Méthode d'Euler

5.2 Méthode d'Euler Modifiée

5.3 Méthode de Runge –Kutta d'ordre 4

Chapitre 6 : Méthode itératives pour les systèmes d'équations linéaires

6.1 Méthode de Jacobi

Cours préparé par
MEZIANE.Assia

Chapitre I : Résolution numérique des équations non linéaires à une variable

Dans ce chapitre on va étudier trois méthodes pour la résolution des équations non linéaires à une variable aussi dites équations à variables transcendentes. Comme exemple de ces équations, on peut citer $\sin(x)+x=0$, $\ln(x)-2x+3=0$. Ces équations ne possèdent pas une ou des racines exactes qui peuvent être calculées directement, c'est pourquoi on fait recours aux méthodes numériques pour trouver les solutions approchées de ces équations. Les racines calculées sont autant précises que l'on veut surtout lorsqu'on dispose des moyens de calculs. Ces méthodes numériques permettent seulement le calcul d'une seule racine sur un intervalle bien choisi. Donc si l'équation possède plus d'une racine, il est nécessaire de les localiser dans des intervalles choisis soigneusement.

1 Localisation des racines d'une équation $f(x)=0$.

Soit une équation $f(x)=0$ dont on cherche la solution sur un intervalle $[a,b]$, on commence par un tracé grossier de la fonction sur l'intervalle donnée puis on isole chaque racine dans un sous intervalle le plus étroit possible. La figure 1 montre le tracé d'une fonction f qui coupe l'axe des x en trois points, c'est-à-dire que l'équation $f(x)=0$ possède trois racines, on note les racines exactes par \bar{x}_1, \bar{x}_2 et \bar{x}_3 .

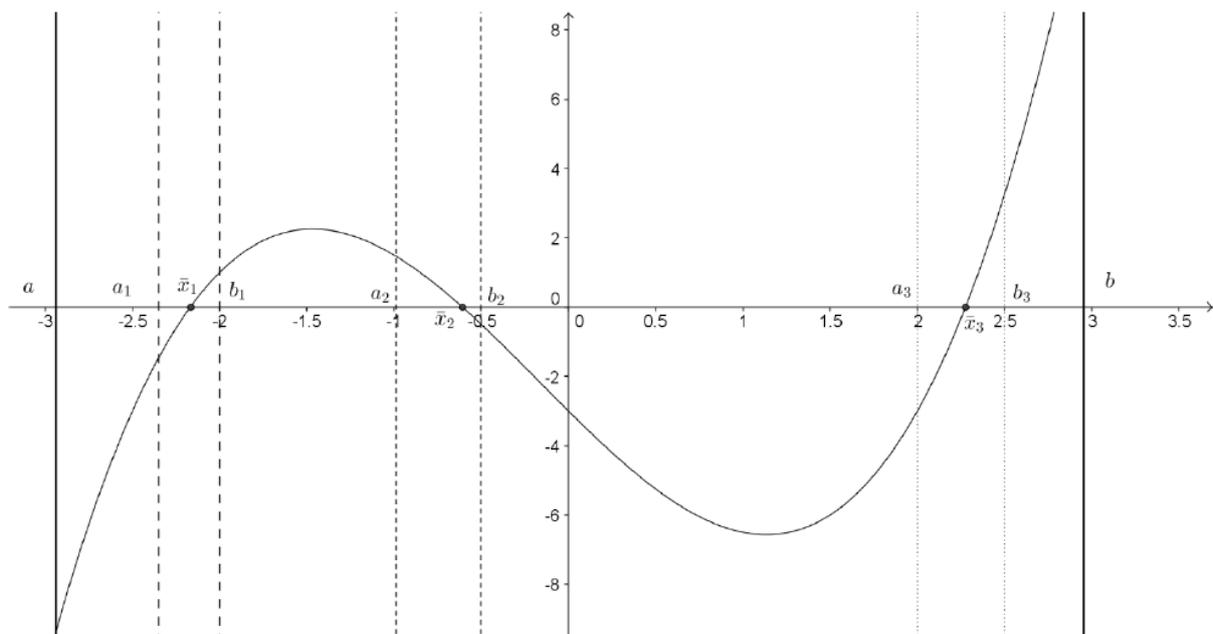


Figure 1 : Localisation des racines

On remarque que la fonction est continue sur chaque sous intervalle, aussi chaque sous intervalle :

Englobe une seule racine tel que $\bar{x}_1 \in [a_1, b_1]$, $\bar{x}_2 \in [a_2, b_2]$ et $\bar{x}_3 \in [a_2, b_3]$.

Vérifie la condition $f(ai)f(bi) < 0$ avec $i=1,2,3$.

La forme de l'équation $f(x)=0$ peut être compliquée, dans ce cas s'il est possible on peut la décomposer en deux parties simples $g(x)=h(x)$.

❖ Par exemple $f(x)=\ln(x)-x^2+2=0$, qui est assez compliqué pour le tracé, peut être décomposée en :

$$g(x)=h(x) \text{ avec } g(x)=\ln(x) \text{ et } h(x)=x^2-2$$

Le tracé de g et h est très simple, la solution de $f(x)=0$ se situe à l'intersection de g et h . Ensuite, on projette les points des intersection et on localise les racines. On peut facilement vérifier les intervalles trouvées en calculant $f(ai)f(bi) < 0$.

2 Méthode de Bisection où de Dichotomie :

C'est la méthode la plus simple et qui nécessite le plus de calculs, elle est basée sur le cas particulier (Théorème de Bolzano) du théorème des valeurs intermédiaires qui dit que :

1. Si $f(x)$ est continue sur l'intervalle $[a, b]$,
2. Si $f(a)$ et $f(b)$ ne sont pas de même signe, il existe au moins un réel c compris entre a et b tel que $f(c) = 0$ (car 0 est compris entre $f(a)$ et $f(b)$).

2.1 Principe de la méthode

Une fois les racines sont localisées chacune dans un intervalle, pour simplifier l'écriture soit par exemple $[a, b]$:

1. On divise l'intervalle en deux parties égales tel que $x_0 = \frac{a+b}{2}$.
2. On obtient deux sous intervalles $[a, x_0]$ et $[x_0, b]$, la racine \bar{x} doit obligatoirement appartenir à l'un d'eux. Pour vérifier, on calcule $f(a)f(x_0)$ et $f(x_0)f(b)$ le produit qui est négatif c'est celui qui correspond à l'intervalle qui contient la solution.

Notons le nouvel intervalle par $[a_1, b_1]$ tel que :

$$a_1 = \begin{cases} a & \text{si } \bar{x} \in [a, x_0] \\ x_0 & \text{si } \bar{x} \in [x_0, b] \end{cases} \quad \text{et} \quad b_1 = \begin{cases} x_0 & \text{si } \bar{x} \in [a, x_0] \\ b & \text{si } \bar{x} \in [x_0, b] \end{cases}$$

En répétant (itérant) la même méthode pour l'intervalle obtenu on aura les valeurs :

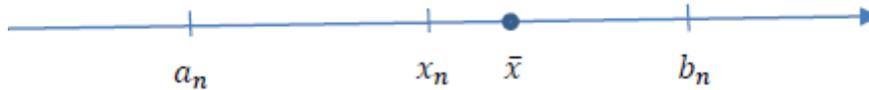
$$x_1 = \frac{a_1+b_1}{2}, x_2 = \frac{a_2+b_2}{2}, \dots, \dots, x_n = \frac{a_n+b_n}{2}$$

La suite $\{x_n\}_{n=0, \infty}$ converge vers la solution \bar{x} de $f(x)=0$.

2.2 Nombre de divisions pour avoir une précision ε donné.

Puisque chaque fois on divise l'intervalle en deux parties égales, on a :

$$\begin{aligned}b_1 - a_1 &= \frac{b - a}{2}; \\b_2 - a_2 &= \frac{b_1 - a_1}{2} = \frac{1}{2} \frac{b - a}{2} = \frac{b - a}{2^2}; \\b_3 - a_3 &= \frac{b_2 - a_2}{2} = \frac{1}{2} \frac{b - a}{2^2} = \frac{b - a}{2^3}; \\&\dots\dots\dots \\b_n - a_n &= \frac{b_{n-1} - a_{n-1}}{2} = \frac{1}{2} \frac{b - a}{2^{n-1}} = \frac{b - a}{2^n}.\end{aligned}$$



Puisque

$$\bar{x} \in [a_n, b_n] = [a_n, x_n] \cup [x_n, b_n] \text{ on a } |x_n - \bar{x}| \leq \frac{1}{2} \frac{b - a}{2^n} = \frac{b - a}{2^{n+1}}$$

Il faut que la différence $|x_n - \bar{x}|$ qui est l'erreur du calcul soit inférieure à une précision donnée ε , c'est-à-dire :

$$|x_n - \bar{x}| \leq \varepsilon$$

Alors, il suffit que :

$$\frac{b - a}{2^{n+1}} \leq \varepsilon$$

Cela donne :

$$n \geq \frac{\ln(\frac{b-a}{2\varepsilon})}{\ln(2)}$$

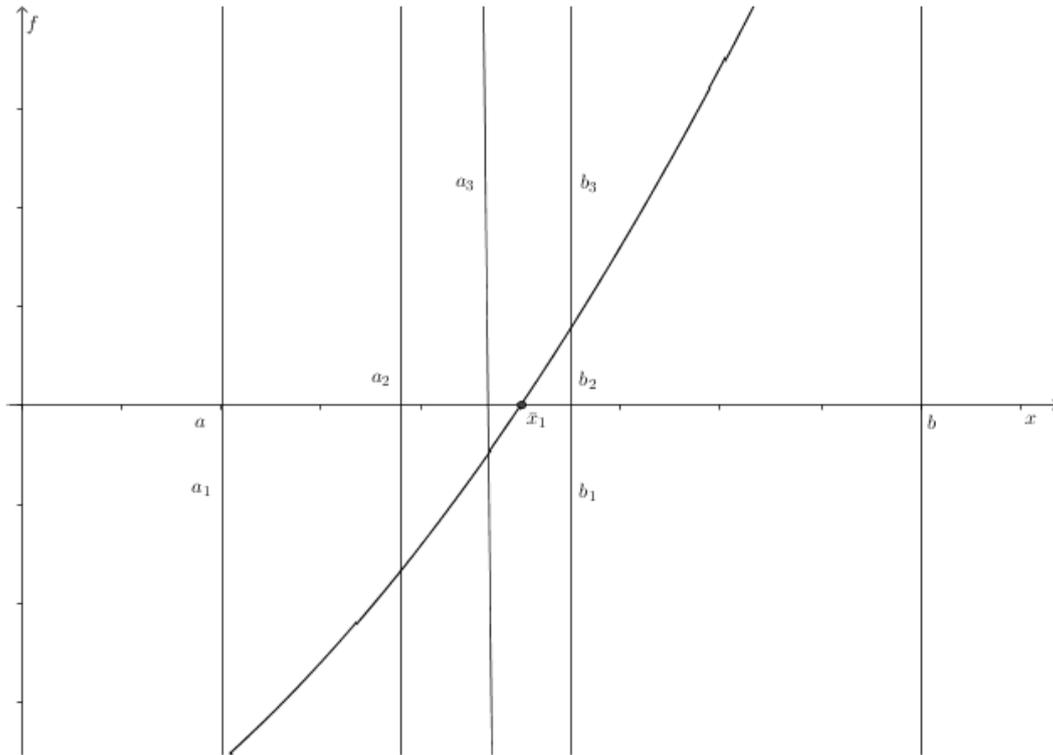


Figure 2 : Illustration de la méthode de bisection

Cette méthode est inconditionnellement convergente, son problème c'est qu'elle est lente c'est pourquoi elle est utilisée pour démarrer d'autres méthodes plus élaborées.

3 Méthode des approximations successives où du point fixe.

Soit g une fonction définie sur un intervalle $[a, b]$, le point \bar{x} qui vérifie $\bar{x} = g(\bar{x})$ avec $\bar{x} \in [a, b]$ est dit point fixe de la fonction g .

Cette méthode est basée sur le principe du point fixe d'une fonction, on écrit l'équation $f(x) = 0$ sous la forme $x = g(x)$, ensuite on cherche le point fixe \bar{x} de la fonction g .

Pour cela on crée la suite $x_{n+1} = g(x_n)$ ($n = 0, 1, 2, \dots$) avec x_0 donnée par dichotomie par exemple.

On part de x_0 pour $n = 0$, on calcule $x_1 = g(x_0)$ ensuite $n = 1$, on calcule $x_2 = g(x_1), \dots, x_{n+1} = g(x_n)$. Sous certaines conditions, la suite $\{x_n\}_{n=0, \infty}$ converge vers la solution \bar{x} point fixe de g et solution de l'équation $f(x) = 0$.

Exemple : Ecrire l'équation $f(x) = 0$ sous la forme $x = g(x)$ si $f(x) = x^2 + 3e^x - 12$.

On peut écrire :

$$x = g_1(x) = x^2 + 3e^x - 12 + x$$

$$x = g_2(x) = \sqrt{12 - 3e^x}$$

$$x = g_3(x) = \ln\left(\frac{12-x^2}{3}\right)$$

Pour pouvoir choisir la forme de g adéquate pour le calcul, un critère de convergence de cette méthode doit être vérifié.

3.1 Critère de convergence de la méthode des approximations successives.

Soit g une fonction dérivable définie de $[a, b] \rightarrow [a, b]$ tel que (condition suffisante) :

$$|g'(x)| \leq k < 1 \quad \forall x \in [a, b]$$

Alors la suite $\{x_n\}_{n=0, \infty}$ définie par $x_{n+1} = g(x_n)$ ($n=0, 1, 2, \dots$) converge indépendamment de la valeur de x_0 vers l'unique point fixe \bar{x} de g .

Si plusieurs formes de g vérifient cette condition, on aura plusieurs valeurs de k . On choisit celle avec la valeur minimale de k . En pratique, on calcule $k = \max_{x \in [a, b]} |g'(x)|$ qui doit être inférieure à l'unité pour que la méthode converge.

4 Méthode de Newton-Raphson :

C'est la méthode la plus efficace et la plus utilisée, elle repose sur le développement de Taylor. Si $f(x)$ est continue et continument dérivable dans le voisinage de \bar{x} solution de $f(x)=0$, alors le développement en série de Taylor autour d'un estimé x_n proche de \bar{x} s'écrit :

$$f(\bar{x}) = f(x_n) + \frac{(\bar{x}-x_n)}{1!} f'(x_n) + \frac{(\bar{x}-x_n)^2}{2!} f''(x_n) + \dots$$

Si x_n est un estimé proche de \bar{x} , alors le carré de l'erreur $\varepsilon_n = \bar{x} - x_n$ et les termes de degrés supérieur sont négligeable. Sachant que $f(\bar{x})=0$ on obtient la relation approximative :

$$f(x_n) + (\bar{x} - x_n)f'(x_n) \approx 0$$

Donc :

$$\bar{x} = x_n - \frac{f(x_n)}{f'(x_n)}$$

On peut écrire la **(n+1)eme** itération approximant \bar{x} est :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (n=0,1,2,\dots)$$

Cette suite, si elle converge, doit converger vers la solution \bar{x} de $f(x)=0$. On remarque que $f'(x)$ doit être non nulle.

4.1 Critère de convergence de la méthode de Newton-Raphson :

Soit une fonction f définie sur $[a, b]$ telle que :

i. $f(a)f(b)<0$

ii. $f'(x)$ et $f''(x)$ sont non nulles et gardent un signe constant sur l'intervalle donné.

4.2 Critère d'arrêt de calcul pour la méthode des approximations successives et Newton-Raphson :

Si la condition de convergence est vérifiée, le procédé itératif doit converger. Cela veut dire que chaque nouvelle itération est meilleure que la précédente, de ce fait on peut dire que si on a une précision ϵ , on arrête le calcul lorsque la différence absolue entre deux approximations successives est inférieure à la précision donnée. C'est-à-dire :

$$|x_{n+1} - x_n| \leq \epsilon$$

Si cette condition est vérifiée on prend x_{n+1} comme solution de $f(x)=0$.

4.1 Critère de convergence de la méthode de Newton-Raphson :

Soit une fonction f définie sur $[a,b]$ telle que :

i. $f(a)f(b) < 0$

ii. $f'(x)$ et $f''(x)$ sont non nulles et gardent un signe constant sur l'intervalle donné.

4.2 Critère d'arrêt de calcul pour la méthode des approximations successives et Newton-Raphson

Si la condition de convergence est vérifiée, le procédé itératif doit converger. Cela veut dire que chaque nouvelle itération est meilleure que la précédente, de ce fait on peut dire que si on a une précision ε , on arrête le calcul lorsque la différence absolue entre deux approximations successives est inférieure à la précision donnée. C'est-à-dire :

$$|x_{n+1} - x_n| \leq \varepsilon$$

Si cette condition est vérifiée on prend x_{n+1} comme solution de $f(x)=0$.

Chapitre II : Interpolation polynômiale

Soit par exemple une expérience où on enregistre la distance parcourue par un objet en fonction du temps, les résultats sont donnés dans le tableau suivant :

t(sec)	0	1	2	3	4
x(m)	0	5	15	0	3

On veut par exemple calculer la position de l'objet au temps $t=2.5$ ou la vitesse de l'objet à un temps donné. Pour cela, il faut avoir une forme analytique de x en fonction de t , $x(t)$. Cette forme doit au moins coïncider avec les points donnés dans le tableau. Ensuite on peut calculer $x(2.5)$,

$$\int_0^4 x(x) dt \text{ où bien } v(t) = \frac{dx(t)}{dt}.$$

Dans ce chapitre, on va considérer l'approximation de $x(t)$ par une forme polynômiale c'est-à-dire :

$$x(x) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$$

Avec a_i ($i=0,n$) sont des coefficients à déterminer.

Les polynômes que nous allons étudier diffèrent seulement par la façon de déterminer les coefficients a_i ($i=0,n$), car pour un tableau de valeurs données le polynôme d'interpolation est unique.

2.1 Polynôme d'interpolation de Lagrange.

Soient $(n+1)$ points distincts x_0, x_1, \dots, x_n et f une fonction dont les valeurs sont $f(x_0), f(x_1), \dots, f(x_n)$. Alors, il existe un seul polynôme de degré inférieur ou égal à n et qui coïncide avec les points d'interpolation, i e :

$$f(x_k) = P_n(x_k), \quad k = 0, 1, 2, \dots, n$$

Ce polynôme est donné par :

$$P_n(x) = \sum_{i=0}^n f(x_i) L_i(x) = f(x_0) L_0(x) + f(x_1) L_1(x) + \dots + f(x_n) L_n(x)$$

Avec :

$$L_k(x) = \sum_{i=0, i \neq k}^n \frac{(x-x_i)}{(x_k-x_i)} = \frac{(x-x_0)}{(x_k-x_0)} \frac{(x-x_1)}{(x_k-x_1)} \dots \frac{(x-x_{k-1})}{(x_k-x_{k-1})} \frac{(x-x_{k+1})}{(x_k-x_{k+1})} \dots \frac{(x-x_n)}{(x_k-x_n)} \quad (k=0, \dots, n)$$

Avec :

$$f[x_0, x_1] = \frac{f[x_1] - f[x_0]}{x_1 - x_0}, \quad f[x_1, x_2] = \frac{f[x_2] - f[x_1]}{x_2 - x_1}, \dots \dots$$

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}, \quad f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1}, \dots \dots$$

$$f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}, \quad f[x_1, x_2, x_3, x_4] = \frac{f[x_2, x_3, x_4] - f[x_1, x_2, x_3]}{x_4 - x_1}$$

$$f[x_0, x_1, x_2, x_3, x_4] = \frac{f[x_1, x_2, x_3, x_4] - f[x_0, x_1, x_2, x_3]}{x_4 - x_0}$$

2.3 Erreur d'interpolation

C'est l'erreur commise lorsqu'on remplace la fonction par le polynôme d'interpolation équivalent. Elle est notée par $\varepsilon(x)$ car elle varie d'un point à un autre dans l'intervalle d'interpolation. Cette erreur doit être nulle aux points d'interpolation, $\varepsilon(x_i) = 0$, ($i=0, \dots, n$).

Si la fonction f est continue et $(n+1)$ fois dérivable sur l'intervalle d'interpolation $[a=x_0, b=x_n]$. Alors pour tout $x \in [a, b]$ il existe $z \in [a, b]$ tel que :

$$\varepsilon(x) = |f(x) - P_n(x)| = \prod_{i=0}^n \frac{(x - x_i)}{(n+1)!} f^{(n+1)}(z)$$

Si $|f^{(n+1)}(z)| \leq M \quad \forall z \in [a, b]$ on peut écrire :

$$\varepsilon(x) = |f(x) - P_n(x)| \leq \prod_{i=0}^n \frac{(x - x_i)}{(n+1)!} M$$

Dans ce cas M est majorant de la fonction $f^{(n+1)}(x)$ sur l'intervalle $[a, b]$.

Chapitre III : Intégration Numérique

Dans ce chapitre, on va étudier quelques méthodes approximatives des intégrales limitées. Aussi ces méthodes permettent le calcul des intégrales qui non pas de solutions directes ou analytiques. On peut aussi calculer l'intégrale d'une fonction donnée sous forme tabulaire ou discrète.

3.1 Formule du trapèze

Cette formule est très simple, elle permet de remplacer la courbe $f(x)$ de la fonction à intégrer par :

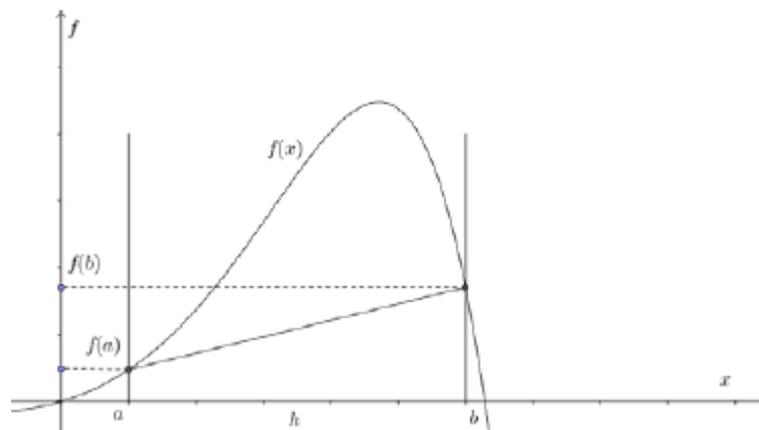


Figure 3 : Méthode du trapèze

une ligne droite qui relie les points $(a, f(a))$ et $(b, f(b))$ ce qui donne un trapèze (voir figure). L'intégrale est donc remplacée par la surface du trapèze :

$$s = \int_a^b f(x) \approx \frac{h}{2} (f(a) + f(b))$$

Avec $h=b-a$ est le pas d'intégration. On peut remarquer qu'il y'a une différence importante entre la courbe de la fonction et la ligne droite, cela veut qu'on commit une erreur de calcul. Pour minimiser cette erreur, on utilise une autre forme plus adaptée de cette formule.

3.1.1 Formule du trapèze généralisée.

On divise l'intervalle $[a, b]$ en plusieurs sous intervalles égaux et on applique la formule du trapèze à chaque sous intervalle. On a donc les sous intervalles $[a=x_0, x_1] \cup [x_1, x_2] \cup \dots \cup [x_{n-1}, x_n=b]$, l'application de la formule du trapèze donne :

$$\int_a^b f(x) = \frac{h}{2}(f(x_0) + f(x_1)) + \frac{h}{2}(f(x_1) + f(x_2)) + \frac{h}{2}(f(x_2) + f(x_3)) + \dots + \frac{h}{2}(f(x_{n-1}) + f(x_n))$$

$$\int_a^b f(x) = \frac{h}{2} \left(f(x_0) + 2 \sum_{i=1}^{n-1} f(x_i) + f(x_n) \right) = \frac{h}{2} \left(f_0 + 2 \sum_{i=1}^{n-1} f_i + f_n \right)$$

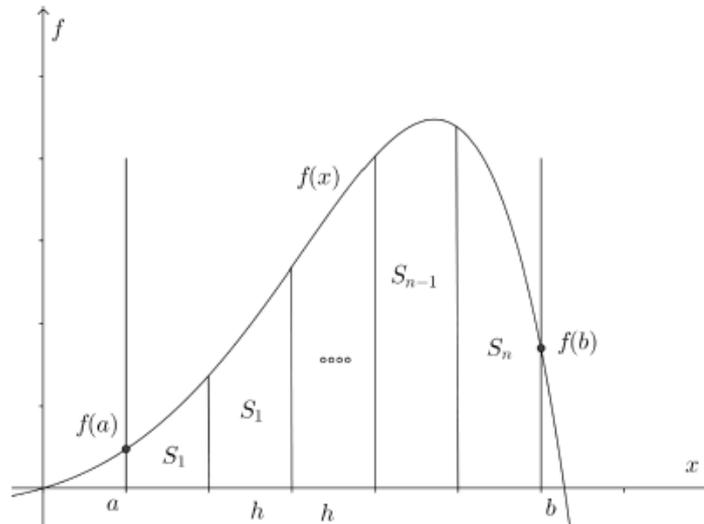


Figure 4 : Méthode du trapèze généralisée

3.1.2 Erreur d'intégration

C'est la différence entre l'intégrale exacte de la fonction et celle calculée par la méthode du trapèze, elle est notée par $R(f)$.

$$R(f) = \int_{x_0}^{x_n} f(x) dx - \frac{h}{2} (f_0 + 2 \sum_{i=1}^{n-1} f_i + f_n) = -\frac{b-a}{12} h^2 f''(z) \text{ avec } z \in [a, b]$$

3.2 Formules de Simpson :

Dans cette formule on ne remplace pas la fonction par une droite mais par une parabole de degré n inférieure ou égale à deux. Cette dernière doit passer par trois points $(x_0, y_0), (x_1, y_1)$ et (x_2, y_2) ce qui fait que cette méthode n'est applicable que pour un nombre pair de tranches (une tranche c'est l'intervalle entre deux points).

La formule de Simpson s'écrit :

$$\int_a^b f(x) \cong \frac{h}{3} (f(x_0) + 4f(x_1) + f(x_2))$$

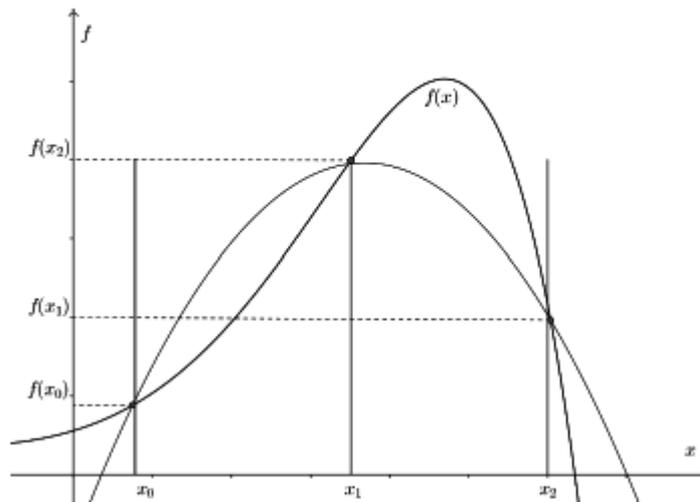


Figure 5 : Méthode de Simpson

Si on généralise la formule de Simpson pour $2n$ sous intervalles avec un pas d'intégration $h = \frac{b-a}{2n}$, $a=x_0 < x_1 < \dots < x_{2n}=b$ et $x_k = a + hk$ pour $k=0, 1, 2, \dots, 2n$. La formule de Simpson généralisée s'écrit :

$$\int_a^b f(x) \cong \frac{h}{3} \left(f(x_0) + 2 \sum_{i \text{ pair}} f(x_i) + 4 \sum_{i \text{ impair}} f(x_i) + f(x_{2n}) \right)$$

L'erreur d'interpolation de la formule de Simpson s'écrit :

$$R(f) = -\frac{b-a}{180} h^4 f^{(4)}(z) \text{ avec } z \in [a, b]$$

3.3 Formules de quadratures :

Dans ces formules on remplace la fonction par un polynôme de Lagrange puis on intègre le polynôme trouvé, on a donc :

$$f(x) \cong P_n(x) = \sum_{i=0}^n f_i L_i(x)$$

Alors :

$$f(x) \cong P_n(x) = \sum_{i=0}^n L_i(x) f_i$$

$$\int_a^b f(x) dx \cong \int_a^b P_n(x) dx = \sum_{i=0}^n f_i \int_a^b L_i(x) dx$$

Si on pose :

$$A_i = \int_a^b L_i(x) dx \text{ pour } i=0,1,\dots,n$$

On obtient la formule de quadrature :

$$\int_a^b f(x) dx \cong \sum_{i=0}^n f_i A_i$$

On doit maintenant choisir la forme de la fonction $f(x)$, dans notre cas on prend :

$$f(x) = x^k \text{ avec } k=0,1,2,\dots,n$$

On remplace dans l'intégrale :

$$\int_a^b x^k dx \cong \sum_{i=0}^n f_i A_i = \frac{b^{k+1}-a^{k+1}}{k+1} \quad k=0,1,2,\dots,n$$

En variant i de 0 à n pour chaque valeur de k on aura pour $k=1$:

$$f_0 A_0 + f_1 A_1 + \dots + f_n A_n = \frac{b^1-a^1}{1} \quad \text{et } f_i = x_i^1$$

D'où on obtient pour toutes les valeurs de i et k le système suivant :

$$\begin{bmatrix} 1 & 1 & 1 & \dots & \dots & 1 \\ x_0 & x_1 & x_2 & \dots & \dots & x_n \\ x_0^2 & x_1^2 & x_2^2 & \dots & \dots & x_n^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_0^n & x_1^n & x_2^n & \dots & \dots & x_n^n \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ \cdot \\ A_n \end{bmatrix} = \begin{bmatrix} I_0 \\ I_1 \\ I_2 \\ \cdot \\ I_n \end{bmatrix} \quad \text{avec } I_k = \frac{b^{k+1}-a^{k+1}}{k+1}$$

Le déterminant de la matrice du système trouvé est dit de "Von Dermonde" il est non nul d'où la solution de ce système existe et unique (A_0, A_1, \dots, A_n) .

On remplace x_n dans l'équation n-1 pour calculer $x_{n-1} = (y_{n-1} - u_{n-1n}x_n) / u_{n-1n-1}$
 Pour le calcul de x_i , les composantes $x_n, x_{n-1}, x_{n-2}, \dots, x_{i+1}$ sont connues, on les remplace dans l'équation i, ce qui donne :

$$x_i = (y_i - \sum_{j=i+1}^n u_{ij}x_j) / u_{ii} \quad i=n-1, n-2, \dots, 1.$$

Le déterminant d'une matrice triangulaire (supérieure ou inférieure) est le produit des éléments de la diagonale (du pivot) de cette matrice :

$$\det(U) = \prod_{i=1}^n u_{ii} = u_{11}u_{22} \dots u_{nn}$$

4.1 Méthode d'élimination de GAUSS

Dans ce chapitre, on va commencer par la méthode d'élimination de GAUSS qui est similaire à celle de l'élimination avec une modification très importante qui facilite énormément la résolution même de systèmes à grandes tailles (des milliers où même des millions). Cette modification transforme un système à matrice A pleine, en un autre à système à matrice U triangulaire supérieure, de telle façon que les deux systèmes $AX=B$ et $UX=Y$ soient équivalents c. à. d. ont la même solution.

L'algèbre linéaire montre que certaines transformations apportées aux systèmes d'équations ne changent pas leurs solutions, dans notre cas les opérations qu'on va appliquer sont les suivantes :

- Multiplication de l'équation E_i par une constante α non nulle, la nouvelle équation obtenue $E_{in} = \alpha E_i$ remplacera l'ancienne E_i .
- Multiplication de l'équation E_j par α non nulle et son ajout à E_i , $E_{in} = E_i + \alpha E_j$, l'équation obtenue E_{in} remplacera E_i .
- Permutation des équations E_i et E_j .

L'application d'une série de ces opérations transformera le système $AX=B$ en $UX=Y$ puis une substitution en arrière donnera la solution du système.

4.1.1 Description de la méthode d'élimination de GAUSS :

On va montrer comment appliquer les transformations au système $\mathbf{AX}=\mathbf{B}$, pour cela le second membre \mathbf{B} sera considéré comme la colonne $(\mathbf{n}+1)$ et sera aussi affecté par les opérations. On divise le travail en $(\mathbf{n}-1)$ étapes chacune d'elle annule les éléments au-dessous du pivot de la colonne $(\mathbf{a}_{ij}$ pour $i>j$). Au début chaque étape, on vérifie que le pivot est non nul. Pour l'étape i le pivot est $a_{ii}^{(i-1)}$ à l'étape $(i-1)$.

Le système à l'état initial où à l'étape (0) est donné par :

$$\begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots a_{nn} \end{bmatrix}^{(0)} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}^{(0)} = \begin{bmatrix} a_{1n+1} \\ a_{2n+1} \\ \vdots \\ a_{nn+1} \end{bmatrix}^{(0)}$$

Première étape :

On vérifie tout d'abord que le pivot de la première étape qui est $a_{ii}^{(i-1)} \neq 0$ pour $i=1$ i.e $a_{11}^{(0)} \neq 0$.

Pour annuler l'élément $a_{21}^{(0)}$ de la deuxième ligne, on multiplie la première équation par $a_{21}^{(0)}$ et on la divise par $a_{11}^{(0)}$ puis on fait la différence de cette nouvelle équation avec la deuxième. L'équation obtenue remplacera la deuxième.

$$E_2^{(1)} = E_2^{(0)} - E_1^{(0)} \frac{a_{21}^{(0)}}{a_{11}^{(0)}}$$

Cette opération donne :

$$a_{21}^{(1)} = 0, a_{22}^{(1)} = a_{22}^{(0)} - a_{12}^{(0)} \frac{a_{21}^{(0)}}{a_{11}^{(0)}}, \dots, a_{2n+1}^{(1)} = a_{2n+1}^{(0)} - a_{1n+1}^{(0)} \frac{a_{21}^{(0)}}{a_{11}^{(0)}}$$

En général, on écrit :

$$a_{2j}^{(1)} = a_{2j}^{(0)} - a_{1j}^{(0)} \frac{a_{21}^{(0)}}{a_{11}^{(0)}} \quad \text{Pour } j = 2, n+1$$

On continue cette procédure avec les lignes 3,4,... Pour la ligne i on a :

$$E_i^{(1)} = E_i^{(0)} - E_1^{(0)} \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}$$

Cette opération donne :

$$a_{i1}^{(1)} = 0, a_{i2}^{(1)} = a_{i2}^{(0)} - a_{12}^{(0)} \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}, \dots, a_{in+1}^{(1)} = a_{in+1}^{(0)} - a_{1n+1}^{(0)} \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}$$

En général, on écrit :

$$a_{ij}^{(1)} = a_{ij}^{(0)} - a_{1j}^{(0)} \frac{a_{i1}^{(0)}}{a_{11}^{(0)}} \quad \text{Pour } i = 2, n \text{ et } j = 2, n + 1$$

A la fin de la première étape, on obtient des éléments nuls au-dessous du pivot de la première étape. Le système s'écrit :

$$\begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & a_{2n}^{(1)} \\ & \vdots & \\ 0 & a_{n2}^{(1)} & a_{nn}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1n+1}^{(0)} \\ a_{2n+1}^{(1)} \\ \vdots \\ a_{nn+1}^{(1)} \end{bmatrix}$$

Deuxième étape :

$$a_{ii}^{(i-1)} \neq 0 \text{ pour } i = 2 \text{ i.e. } a_{22}^{(1)} \neq 0.$$

De la même façon, on obtient pour le cas général :

$$a_{ij}^{(2)} = a_{ij}^{(1)} - a_{2j}^{(1)} \frac{a_{i2}^{(1)}}{a_{22}^{(1)}} \quad \text{Pour } i = 3, n \text{ et } j = 3, n + 1$$

Étape k :

$$a_{ii}^{(i-1)} \neq 0 \text{ pour } i = k \text{ i.e. } a_{kk}^{(k-1)} \neq 0.$$

Pour une étape k quelconque, on a :

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{kj}^{(k-1)} \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \quad \text{Pour } k = \overline{1, n-1}, \quad i = \overline{k+1, n} \text{ et } j = \overline{k+1, n+1}$$

A la fin de la procédure, on obtient un système à matrice triangulaire supérieure qui s'écrit :

$$\begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & a_{2n}^{(1)} \\ & \vdots & \\ 0 & 0 & a_{nn}^{(n-1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1n+1}^{(0)} \\ a_{2n+1}^{(1)} \\ \vdots \\ a_{nn+1}^{(n-1)} \end{bmatrix}$$

La résolution de ce système se fait par substitution en arrière.

4.1.2 Le nombre d'opérations nécessaires pour l'application de l'algorithme de GAUSS est :

- **Nombre de multiplications et d'additions :**

$$nm = na = \frac{n(n-1)(2n+5)}{6}$$

- **Nombre de divisions :**

$$nd = \frac{n(n+1)}{2}$$

4.1.3 Applications de la méthode de GAUSS.

Calcul du déterminant d'une matrice triangulaire :

Le déterminant d'une matrice triangulaire est donné par :

$$\det(U) = (-1)^P \prod_{i=1}^n u_{ii} = (-1)^P u_{11} u_{22} \dots u_{nn}$$

Avec P le nombre de permutation de lignes ou de colonnes effectuées lors de l'application de l'algorithme de GAUSS.

Résolution simultanée de plusieurs systèmes :

Dans la pratique, on rencontre souvent le cas de plusieurs systèmes d'équations qui ne diffèrent que par le second membre B. On peut appliquer l'algorithme de GAUSS sur la matrice A augmentée par tous les seconds membres. De cette façon, on fait les calculs une seule fois sur la matrice A, la substitution se fait avec chaque second membre obtenu à part.

$$\begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} b_1 & c_1 & z_1 \\ b_2 & c_2 & \dots & z_2 \\ & \vdots & & \\ b_n & c_n & \dots & z_n \end{bmatrix}$$

Calcul de l'inverse d'une matrice :

Si A est une matrice d'ordre n , la matrice A^{-1} tel que $AA^{-1}=I$ est dite matrice inverse de A .

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots a_{nn} \end{bmatrix} \text{ et } A^{-1} = \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & \dots x_{2n} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots x_{nn} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & \dots x_{2n} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots x_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \dots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \dots 1 \end{bmatrix}$$

5.1.4 Utilisation de la pivotation.

Dans l'élaboration de l'algorithme de GAUSS, on a supposé que le pivot n'est pas nul, ce n'est pas le cas toujours. Parfois le pivot est très petit comparativement aux autres termes ou même nul, dans ce cas on peut utiliser la technique de la pivotation soit partielle ou totale.

Pivotation partielle :

Dans ce cas on choisit comme pivot l'élément $a_{ik}^{(k-1)}$ tel que :

$$a_{ik}^{(k-1)} = \max_{i \in [k, n]} |a_{ik}^{(k-1)}|$$
$$\begin{bmatrix} a_{11} & \dots & a_{1k} & a_{1n} \\ 0 & \dots & a_{kk} & \dots a_{kn} \\ \vdots & & \vdots & \\ 0 & \dots & a_{nk} & \dots a_{nn} \end{bmatrix}$$

Dans la pivotation partielle, on utilise la permutation des lignes ceci n'a aucun effet sur la solution du système.

Pivotation totale :

Dans la pivotation totale le choix du pivot se fait à partir d'une sous matrice incluant la permutation des lignes et des colonnes tel que :

$$a_{lm}^{(k-1)} = \max_{i, j \in [k, n]} |a_{ij}^{(k-1)}|$$
$$\begin{bmatrix} a_{11} & \dots & a_{1k} & a_{1n} \\ 0 & \dots & a_{kk} & \dots a_{kn} \\ \vdots & & \vdots & \\ 0 & \dots & a_{nk} & \dots a_{nn} \end{bmatrix}$$

On a donc $AX=B$ (1) et $A=LU$ donc $LUX=B$, on pose $UX=Y$ (Y vecteur inconnu), cela donne :

$$\begin{cases} LY = B \\ UX = Y \end{cases} \quad (2)$$

Le système (1) est décomposé en deux systèmes triangulaires faciles à résoudre (2). Le système à matrice triangulaire supérieure et résolu par substitution en arrière, celui à matrice triangulaire inférieure par substitution en avant.

4.3.1 Détermination des matrices L et U :

$$L = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n-1,1} & l_{n-1,2} & \dots & l_{n-1,n-1} & 0 \\ l_{n1} & l_{n2} & \dots & l_{nn-1} & l_{nn} \end{bmatrix} \quad \text{et} \quad U = \begin{bmatrix} 1 & u_{12} & \dots & u_{1n-1} & u_{1n} \\ 0 & 1 & \dots & u_{2n-2} & u_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & u_{n-1n} \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

Les éléments de chaque matrice sont donnés par :

$$\begin{cases} l_{ki} = a_{ki} - \sum_{j=1}^{i-1} l_{kj} u_{ji} \\ u_{ik} = [a_{ik} - \sum_{j=1}^{i-1} l_{ij} u_{jk}] / l_{ii} \end{cases} \quad i = 2, 3, \dots, n \text{ et } k = i, i+1, \dots, n$$

En pratique pour calculer les éléments des deux matrices, on divise la tâche en plusieurs étapes. Par exemple dans l'étape i on détermine :

- La colonne i de L en multipliant L par la colonne i de U .
- La ligne i de U en multipliant la ligne i de L par U .

4.4 Méthode de Choleski :

Cette méthode est applicable aux systèmes à matrices symétriques définies positives ($\text{Det}(A) > 0$ et a_{ij} réels). Cherchons une matrice M telle que $A=MM^t$ ou M est triangulaire inférieure et M^t la matrice transposée de M .

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1n-1} & m_{1n} \\ 0 & m_{22} & \dots & m_{2n-2} & m_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & m_{n-1n-1} & m_{n-1n} \\ & & 0 & m_{nn} & 1 \end{bmatrix}$$

Les éléments de la matrice \mathbf{M} sont donnée par :

$$\begin{cases} m_{ii} = \sqrt{a_{ii} - \sum_{j=1}^{i-1} m_{ij}^2} \\ m_{ji} = \left[a_{ij} - \sum_{k=1}^{i-1} m_{ik} m_{jk} \right] / m_{ii} \end{cases} \quad i = \overline{1, n} \text{ et } j = \overline{i+1, n}$$

chapitre V : Résolution numérique des équations différentielles ordinaires : Problème de Cauchy

Les équations différentielles sont utilisées dans la modélisation mathématique de la quasi-totalité des phénomènes physiques. Une équation différentielle est une relation entre une variable et sa ou ses dérivées de divers ordres.

Dans ce chapitre, on va considérer les équations différentielles ordinaires du premier ordre avec une condition initiale imposée (problème de Cauchy).

Le problème de Cauchy est défini par la solution de l'équation différentielle suivante :

$$\begin{cases} \frac{dy(t)}{dt} = y' = f(t, y(t)) \\ y(t_0) = y_0 \end{cases}$$

$y(t_0) = y_0$ est une condition initiale imposée.

On remarque que la solution doit obligatoirement passer par le point (t_0, y_0) .

La solution du problème de Cauchy existe et est unique si la fonction $f(t, y(t))$ satisfait la condition de Lipschitz en y sur un rectangle R défini par $a \leq t \leq b$ et $c \leq y \leq d$. Cette condition nécessite que :

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2|.$$

En pratique pour vérifier cette condition, on calcule :

$$\left| \frac{df(t, y)}{dy} \right| \leq L \text{ sur } R$$

5.1 Méthode d'Euler :

Soit un intervalle $[a, b]$ sur lequel on cherche la solution d'un problème de Cauchy :

$$\begin{cases} \frac{dy(t)}{dt} = y' = f(t, y(t)) \\ y(t_0 = a) = y_0 \end{cases}$$

La fonction $f(t, y(t))$ vérifie la condition de Lipschitz en y sur le rectangle R .

La première étape consiste à diviser l'intervalle donné en n points équidistants ce qui donne un pas d'intégration $h = \frac{b-a}{n}$, le point d'abscisse t_i est donné par $t_i = a + ih$ (pour $i=1, 2, \dots, n$).

On va donc résoudre le problème sur l'intervalle $[a=t_0, b=t_n]$ avec $y(t_0=a)=y_0$. Si les fonctions $y(t), y'(t)$ et $y''(t)$ sont continués, on peut écrire le développement en série de Taylor pour $y(t)$ au voisinage de t_0 .

$$y(t) = y(t_0) + y'(t_0) \frac{(t - t_0)}{1!} + y''(t_0) \frac{(t - t_0)^2}{2!} + \dots$$

Sachant que :

$$y'(t_0) = f(t_0, y(t_0)) \text{ et } h = t_1 - t_0,$$

Ecrivons :

$$y(t_1) = y(t_0) + f(t_0, y(t_0))h + y''(t_0) \frac{h^2}{2!} + \dots$$

Si h est suffisamment petit alors on peut négliger h^2 et donc les termes d'ordre deux et plus, on obtient donc :

$$y(t_1) = y(t_0) + hf(t_0, y(t_0)) \rightarrow y_1 = y_0 + hf(t_0, y_0)$$

C'est l'approximation d'Euler d'ordre 1, en répétant le procédé on génère une séquence de points approximant $y=y(t)$, en général :

$$\begin{cases} t_i = a + ih, & i = 0, 1, 2, \dots, n-1 \\ y_{i+1} = y_i + hf(t_i, y_i), & y_0 = y(t_0) \end{cases}$$

5.2 Méthode d'Euler modifiée

D'après le développement en série de Taylor, on a :

$$y(t) = y(t_0) + y'(t_0) \frac{(t - t_0)}{1!} + y''(t_0) \frac{(t - t_0)^2}{2!} + y'''(t_0) \frac{(t - t_0)^3}{3!} + \dots$$

Si on prend les trois premier termes et on néglige les autres d'ordre supérieur, on obtient :

$$y(t) \cong y(t_0) + y'(t_0) \frac{(t - t_0)}{1!} + y''(t_0) \frac{(t - t_0)^2}{2!} + \dots$$

En remplaçant t par t_1 cela donne :

$$y(t_1) = y(t_0) + y'(t_0) \frac{(t_1 - t_0)}{1!} + y''(t_0) \frac{(t_1 - t_0)^2}{2!} = y(t_0) + y'(t_0)h + y''(t_0) \frac{h^2}{2!}$$

Or :

$$y''(t_0) = \frac{y'(t_1) - y'(t_0)}{h} \text{ on aura } y(t_1) = y(t_0) + \frac{h}{2}[f(t_1, y_1) + f(t_0, y_0)]$$

En général la formule de Newton modifiée s'écrit :

$$y_{i+1} = y_i + \frac{h}{2}[f(t_{i+1}, y_{i+1}^p) + f(t_i, y_i)]$$

On remarque que cette formule donne y_{i+1} en fonction de y_{i+1}^p qui doit être calculée par la méthode d'Euler.

5.3 Méthode de Runge-Kutta d'ordre 4 :

C'est la méthode la plus précise et la plus utilisée, elle est d'ordre quatre. L'intervalle $[a, b]$ est divisé en n sous intervalles de largeur h , cette formule s'écrit :

$$y_{i+1} = y_i + \frac{h}{6}(K_1 + 2(K_2 + K_3) + K_4)$$

$$K_1 = f(t_i, y_i)$$

$$K_2 = f\left(t_i + \frac{h}{2}, y_i + \frac{h}{2}K_1\right)$$

$$K_3 = f\left(t_i + \frac{h}{2}, y_i + \frac{h}{2}K_2\right)$$

$$K_4 = f(t_i + h, y_i + hK_3)$$

