



HAL
open science

Étude des erreurs dans le fonctionnement d'un analyseur différentiel digital

Claude Echevin

► **To cite this version:**

Claude Echevin. Étude des erreurs dans le fonctionnement d'un analyseur différentiel digital. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1964. Français. NNT: . tel-00278849

HAL Id: tel-00278849

<https://theses.hal.science/tel-00278849>

Submitted on 14 May 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre

T H E S E

présentée à la Faculté des Sciences
de l'Université de GRENOBLE

pour obtenir

Le titre de Docteur de Troisième Cycle

" Mathématiques Appliquées "

par

Claude E C H E V I N

ETUDE DES ERREURS DANS LE FONCTIONNEMENT D'UN

ANALYSEUR DIFFERENTIEL DIGITAL

Thèse soutenue le 23 Octobre 1964 devant la Commission d'Examen :

MM. KUNTZMANN

GASTINEL

HACQUES

N° d'ordre

T H E S E

présentée à la Faculté des Sciences
de l'Université de GRENOBLE

pour obtenir

Le titre de Docteur de Troisième Cycle

" Mathématiques Appliquées "

par

Claude E C H E V I N

ETUDE DES ERREURS DANS LE FONCTIONNEMENT D'UN

ANALYSEUR DIFFERENTIEL DIGITAL

Thèse soutenue le 23 Octobre 1964 devant la Commission d'Examen :

MM. KUNTZMANN

GASTINEL

HACQUES

L I S T E D E S P R O F E S S E U R S

DOYENS HONORAIRES

M. FORTRAT P.

M. MORET L.

DOYEN

M. WEIL L.

PROFESSEURS TITULAIRES

MM. NEEL L.	MAGNETISME ET PHYSIQUE DU SOLIDE
DORIER A.	ZOOLOGIE
HEILMANN R.	CHIMIE ORGANIQUE
KRAVTCHENKO J.	MECANIQUE RATIONNELLE
CHABAUTY C.	CALCUL DIFFERENTIEL ET INTEGRAL
PARDE M.	POTAMOLOGIE
BENOIT J.	RADIOELECTRICITE
CHENE M.	CHIMIE PAPETIERE
BESSON J.	ELECTROCHIMIE
WEIL L.	THERMODYNAMIQUE
FELICI N.	ELECTROSTATIQUE
KUNTZMANN J.	MATHEMATIQUES APPLIQUEES
BARBIER R.	GEOLOGIE APPLIQUEE
SANTON L.	MECANIQUE DES FLUIDES
OZENDA P.	BOTANIQUE
FALLOT M.	PHYSIQUE INDUSTRIELLE
GALVANI O.	MATHEMATIQUES
MOUSSA A.	CHIMIE NUCLEAIRE
TRAYNARD P.	CHIMIE
SOUTIF M.	PHYSIQUE
CRAYA A.	HYDRODYNAMIQUE
REULOS R.	THEORIE DES CHAMPS
AYANT Y.	PHYSIQUE APPROFONDIE
GALLISSOT F.	MATHEMATIQUES APPLIQUEES
Melle LUTZ E.	MATHEMATIQUES
MM. BLAMBERT M.	MATHEMATIQUES
BOUCHEZ R.	PHYSIQUE NUCLEAIRE
ILLIBOUTRY L.	GEOPHYSIQUE
MICHEL R.	GEOLOGIE ET MINERALOGIE
BONNIER E.	ELECTROCHIMIE
DESSAUX G.	PHYSIQUE ANIMALE
PILLET E.	ELECTROCHIMIE
DEBELMAS J.	GEOLOGIE
GERBER R.	MATHEMATIQUES
PAUTHENET R.	ELECTROTECHNIQUE
VAUQUOIS B.	MATHEMATIQUES APPLIQUEES
SILBER R.	MECANIQUE DES FLUIDES
MOUSSIEGT J.	ELECTRONIQUE
BARBIER J. C.	PHYSIQUE
KPSZUL J. L.	MATHEMATIQUES
BUYLE-BODIN M.	ELECTRONIQUE

PROFESSEURS SANS CHAIRE

M.	LACASE A.	THERMODYNAMIQUE
Mme	KOFLER L.	BOTANIQUE
MM.	DREYFUS B.	THERMODYNAMIQUE
	VAILLANT F.	ZOOLOGIE ET HYDROBIOLOGIE
	GIRAUD P.	GEOLOGIE
	GIDON P.	GEOLOGIE ET MINERALOGIE
	ARNAUD P.	CHIMIE
	PERRET R.	SERVOMECHANISMES
Mme	LUMER L.	MATHEMATIQUES
Mme	BARBIER M. J.	ELECTROCHIMIE
Mme	SOUTIF J.	PHYSIQUE
MM.	BRISSONNEAU P.	PHYSIQUE
	COHEN J.	ELECTROCHIMIE
	DEPASSEL R.	MECANIQUE
	GASTINEL N.	MATHEMATIQUES APPLIQUEES

PROFESSEURS ASSOCIES

MM.	LUMER G.	MATHEMATIQUES
	HIGUCHI	BIOSYNTHESE DE LA CELLULOSE
	WAGNER	BOTANIQUE

MAITRES DE CONFERENCES

MM.	ROBERT A.	CHIMIE PAPETIERE
	ANGLES D'AURIAC	MECANIQUE DES FLUIDES
	BIAREZ J. P.	MECANIQUE PHYSIQUE
	COUMES A.	ELECTRONIQUE
	DODU J.	MECANIQUE DES FLUIDES
	DUCROS P.	MINERALOGIE ET CRISTALLOGRAPHIE
	CLENAT P.	CHIMIE
	HACQUES G.	CALCUL NUMERIQUE
	LANCIA R.	PHYSIQUE AUTOMATIQUE
	PEBAY-PEROULA	PHYSIQUE
	KAHANE	PHYSIQUE GENERALE
	DOLIQUE	ELECTRONIQUE
Mme	KAHANE J.	PHYSIQUE
MM.	DEGRANGE C.	ZOOLOGIE
	GAGNAIRE D.	CHIMIE PAPETIERE
	RASSAT A.	CHIMIE SYSTEMATIQUE
	KLEIN J.	MATHEMATIQUES
	BETHOUX P.	MATHEMATIQUES APPLIQUEES
	POULOUJADOFF M.	ELECTROTECHNIQUE
	DEPOMMIER P.	PHYSIQUE NUCLEAIRE
	DEPORTES C.	CHIMIE
	BARRA J.	MATHEMATIQUES APPLIQUEES
Mme	BOUCHE L.	MATHEMATIQUES
MM.	PERRIAUX J.	GEOLOGIE
	SARROT-REYNAULD	GEOLOGIE
	CAUQUIS G.	CHIMIE GENERALE
	LABBE A.	BOTANIQUE
	BONNET G.	PHYSIQUE GENERALE
	BARNOUD F.	BIOSYNTHESE DE LA CELLULOSE
Mme	BONNIER M. J.	CHIMIE

MAITRES DE CONFERENCES ASSOCIES

MM.	ISHIKAWA Y.	MAGNETISME
	QUATTROPANI	THERMODYNAMIQUE

Je tiens à exprimer ma profonde reconnaissance.

à

Monsieur le Professeur KUNTZMANN

Directeur du Laboratoire de Calcul de l'Université de GRENOBLE, pour le temps qu'il a consacré à la direction de cette étude et pour l'honneur qu'il m'a fait en acceptant de présider le jury de thèse.

à

Monsieur le Professeur GASTINEL

et à

Monsieur HACQUES Maître de Conférences

qui ont bien voulu faire partie du jury.

Je remercie également :

Monsieur VIVIER pour l'intérêt qu'il a porté à cette étude et pour ses conseils judicieux.

Monsieur BELLINO et Monsieur MARTIN qui m'ont permis d'effectuer les calculs numériques dans de très bonnes conditions.

Madame FRAIMBAULT et Mademoiselle MEUNIER pour la réalisation matérielle de cette thèse.

AVANT - PROPOS

Le présent travail est une étude des erreurs dans le fonctionnement d'un analyseur différentiel digital en cours de construction dans les laboratoires de l'Université de Grenoble.

Le fonctionnement de l'appareil a été simulé sur un ordinateur GAMMA ET Bull ; la description de cette simulation est faite en annexe.

Dans l'introduction, nous rappellerons le principe de l'analyseur différentiel digital et définirons les caractéristiques de l'appareil étudié.

INTRODUCTION

A. - Rappel du principe de l'analyseur différentiel digital (ADD)

A. - 1 Principe de l'intégration.

Etant donnée une fonction $y(t)$ dont on connaît les valeurs $y(t_0), \dots, y(t_i)$ pour des valeurs t_0, \dots, t_i de la variable ($t_i = t_{i-1} + h$), on peut trouver une approximation de la primitive $z(t) = \int_{t_0}^t y(u) du$ par la méthode très simple des

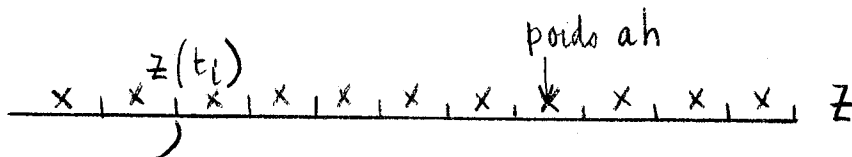
rectangles.

Les valeurs successives de $z(t_i)$ sont données par la relation

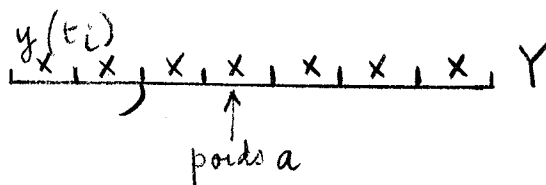
$$z(t_i) = z(t_{i-1}) + hy(t_{i-1})$$

Cette relation est aisément représentable à l'aide de deux registres Y et Z et d'un additionneur.

Si Y et Z contiennent respectivement $y(t_i)$ et $z(t_i)$, si Y est additionné à Z en cadran Y et Z de telle manière qu'à une position de $y(t_i)$ de poids a corresponde une position de poids ah, alors, après l'addition, le registre Z contient $z(t_{i+1})$



ADDITIONNEUR



Supposons que, lorsque t prend successivement les valeurs t_0, t_1, \dots, t_i , le registre Y contienne les valeurs $y(t_0), y(t_1), \dots, y(t_i)$. Pour que le registre Z contienne les valeurs $z(t_0), \dots, z(t_i)$ lorsque $t = t_0, t_1, \dots, t_i$, il suffit d'assurer l'addition du contenu de Y au contenu de Z dans les conditions de cadrage définies plus haut à chaque fois que t passe de la valeur t_j à la valeur t_{j+1} , c'est-à-dire que l'on veut effectuer le pas d'intégration.

Cette opération est effectuée automatiquement par la machine à des intervalles de temps constants.

A. 2 Représentation d'un nombre dans un registre

L'analyseur différentiel étudié utilise la numération binaire. Le poids d'une position d'un registre et le pas d'intégration sont donc des puissances de 2.

On peut toujours considérer que le contenu d'un registre est, en valeur absolue, un nombre inférieur ou égal (cas où le nombre est négatif) à 1.

Dans le paragraphe précédent, le contenu du registre Y pouvait être noté Y_i . Si q est le nombre de positions binaires du registre Y à gauche de la virgule de $y(t_i)$; il est évident que :

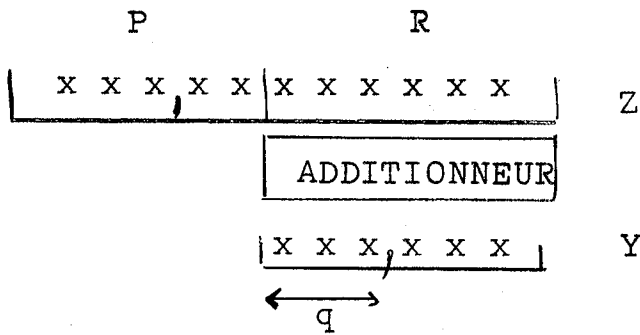
$$y(t_i) = 2^q \cdot Y_i$$

q est appelé facteur de cadrage.

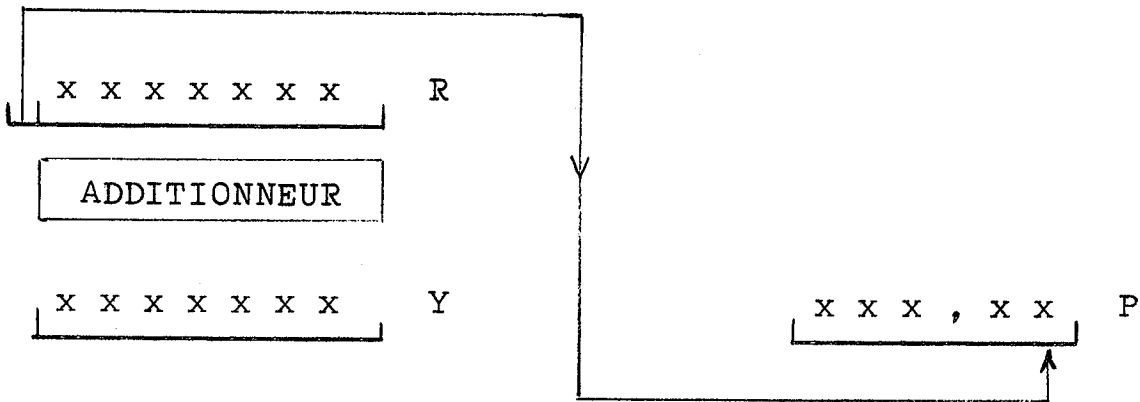
A. 3 Principe de la liaison par incrément.

Le registre Z défini plus haut peut être considéré comme formé de deux parties :

- Une partie R des faibles poids à laquelle est additionné Y ,
- Une partie P des poids forts qui reçoit uniquement les débordements de la partie R et les accumule.



La partie poids forts P joue le rôle d'un compteur. Elle peut être dissociée de la partie poids faibles R du registre Z s'il existe un moyen de transmission d'un débordement de R vers P. Cette transmission est réalisée dans l'analyseur différentiel digital.



Un débordement de R au pas d'intégration i exprime la variation $hy(t_i)$ de la quantité $z(t_i)$ lorsque t passe de la valeur t_i à t_{i+1} .

Dans l'analyseur différentiel digital, la liaison entre les opérateurs a lieu par transmission des variations des variables.

Du fait de l'utilisation de la numération binaire, le débordement de R ne peut prendre que les valeurs 0, 1 ou - 1.

Si a est le poids de la position d'extrême gauche du registre Y, le poids du débordement est $2 a.h$

Un tel débordement de valeur fixe est appelé incrément.

Nous n'insisterons pas davantage sur le principe de l'intégration et de la liaison entre intégrateurs par transmission de débordements.

Lorsque, au cours de l'exposé, nous ferons intervenir des opérateurs différents de l'intégrateur, nous en rappellerons brièvement le principe.

A. 4 Premières notations.

Dans un intégrateur, il n'existe ainsi que deux registres R et Y. Le registre R a une ou deux (cas de l'incrément ternaire) positions de plus que le registre Y.

Le contenu R_i de R au pas i peut être dissocié de la manière suivante :

- Une partie entière $E (R_i)$ qui est l'incrément émis par R
- Une partie fractionnaire $F (R_i)$

$$\boxed{R_i = E (R_i) + F (R_i)}$$

B. - Exemple du problème $y' = - y$

Schéma représentatif, Relations fondamentales, Erreurs.

B. 1 Schéma représentatif

Soit le problème différentiel aux conditions initiales

$$\begin{cases} y' = - y \\ y (t_0) = y_0 \end{cases}$$

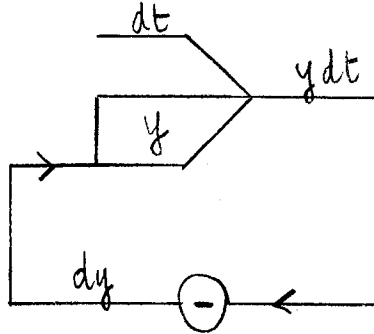
qui équivaut à

$$y(t) - y (t_0) = - \int_{t_0}^t y(u) du$$

Un intégrateur doit être utilisée pour résoudre ce problème.

- $y(t_0)$ doit être mis initialement dans le registre Y
- Les débordements de R, changés de signe, doivent être accumulés dans Y.

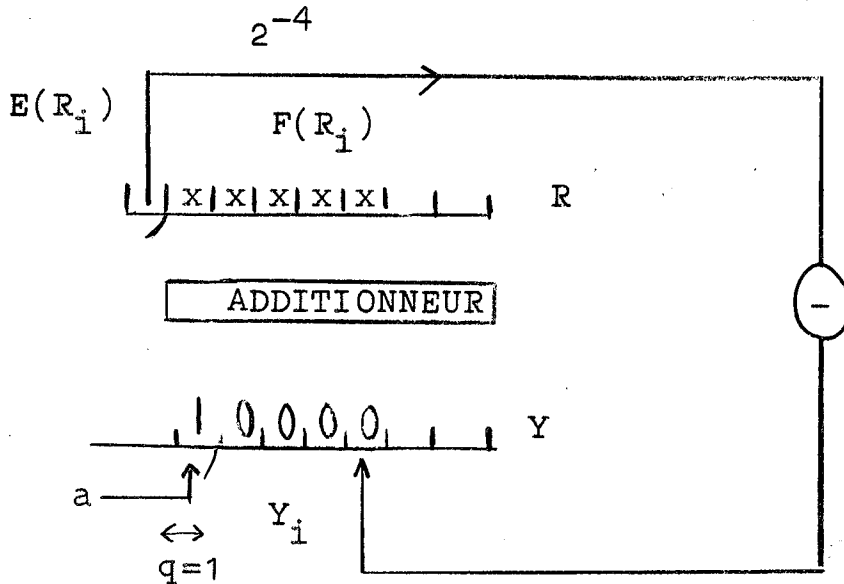
Le schéma de résolution de cette équation différentielle est le plus couramment représenté comme suit :



D'une manière plus précise, supposons que $h = 2^{-5}$

$$y_0 = 1, x_0 = 0, \quad q \text{ (facteur de cadrage)} = 1$$

En utilisant les notations définies plus haut, le schéma se présente comme suit



Avant tout calcul, le registre Y contient un nombre $y_0 = 2^{-q} \cdot y_0(t) = \frac{1}{2}$

Le poids a de la position extrême gauche du registre Y est 1, le poids de l'incrément $E(R_i)$ est donc 2^{-4}

Les débordements de R doivent donc être accumulés sur la position de poids 2^{-4} du registre Y.

B. 2 Relations fondamentales

A partir du schéma précédent, on peut définir des relations entre les diverses variables. Nous appellerons ces relations, relations **fondamentales**

$$\begin{cases} Y_i = Y_{i-1} - hE(R_{i-1}) & E(R_{i-1}) = 0 \text{ ou } 1 \\ R_{i-1} = F(R_{i-2}) + Y_{i-1} \end{cases}$$

Du fait que $R_{i-1} = E(R_{i-1}) + F(R_{i-1})$

L'incrément $E(R_{i-1})$ émis au pas $i - 1$ a la valeur

$$E(R_{i-1}) = Y_{i-1} - [F(R_{i-1}) - F(R_{i-2})]$$

Définissons une variable λ_j égale à $F(R_j) - F(R_{j-1})$

avec $\lambda_0 = F(R_0)$

Y_i peut finalement s'écrire

$$Y_i = Y_{i-1} - h Y_{i-1} + h \lambda_{i-1}$$

B. 3 Erreurs commises dans la résolution de $y' = -y$ par l'ADD

On peut distinguer deux erreurs à causes différentes

B. 3 α - Une erreur de méthode

Soit (y_i, x_i) un point exact de la courbe intégrale étudiée, et (y_i, x_i) le point correspondant obtenu par la méthode d'intégration approchée des rectangles.

La valeur $\sigma_i = y_i - y_i$ est l'erreur de méthode au pas i .

B. 3 β Une erreur de principe (ou systématique)

La valeur y_i est représentée dans le registre Y par Y_i
 $2^q Y_i$ est comparable à y_i .

Nous appellerons $\epsilon_i = 2^q Y_i - y_i$ l'erreur de principe au pas i .

Causes de l'erreur de principe

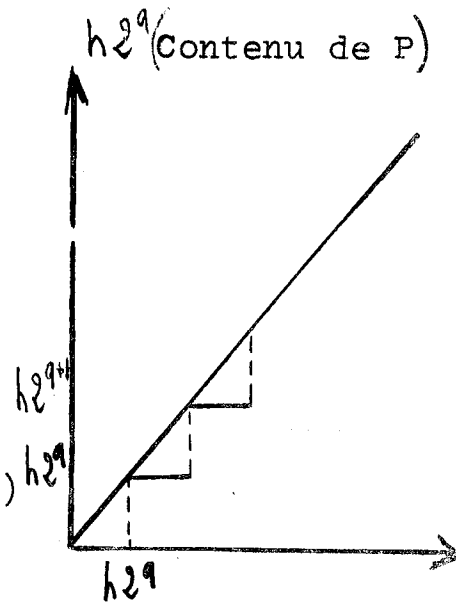
a) Arrondi sur la valeur d'une intégrale

Nous avons vu qu'une intégrale calculée par l'intégrateur était contenue en double longueur dans le registre R (poids faibles) et le registre P (poids forts).

Or seule la partie poids forts de l'intégrale est utilisée.

Au pas d'intégration i , la valeur exacte de l'intégrale est

$$z_i = h 2^q \underbrace{\sum_0^i E(R_k)}_{\text{contenu de P}} + h 2^q \underbrace{F(R_i)}_{\text{contenu de R}}$$



Nous appellerons arrondi au pas i sur la valeur de l'intégrale et nous noterons $a(i)$ la valeur $h 2^q F(R_i)$

L'arrondi peut-être représenté graphiquement

$$a(i) = (\text{contenu de Z} - \text{contenu de P}) h 2^q$$

Contenu de Z

L'arrondi est évidemment une cause d'erreur.

L'erreur sur le contenu d'un registre Y provoquée par des arrondis sera appelée erreur d'arrondi.

b) Retards intervenant entre la production et l'utilisation des incréments.

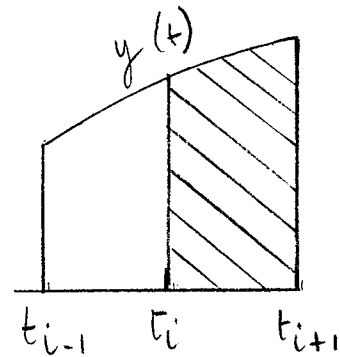
a) Reprenons l'exemple de la résolution de l'équation différentielle $y' = -y$.

Au pas d'intégration i , le contenu du registre Y passe de la valeur Y_{i-1} à la valeur Y_i

L'incrément émis $E(R_i)$ modifie le registre Y au pas suivant $i + 1$. Il y a donc un retard d'un cycle entre la production et l'utilisation de cet incrément. Cet incrément doit représenter la valeur $h y_i$, approximation de l'accroissement

$$\int_{t_i}^{t_{i+1}} y(u) du$$

C'est-à-dire l'aire hachurée de la figure



Or au pas i , la valeur hY_i est additionnée au contenu du registre R . L'incrément $E(R_i)$ émis est donc une bonne représentation de

tion de $\int_{t_i}^{t_{i+1}} y(u) du$.

β) Nous étudions un analyseur différentiel de type séquentiel

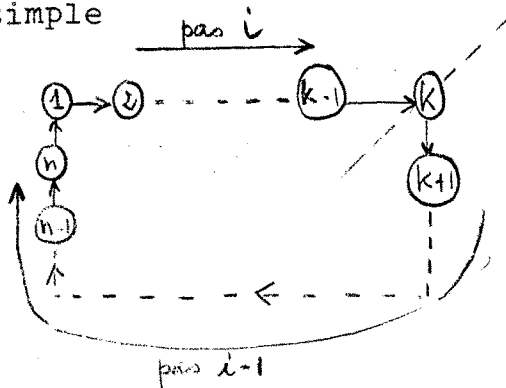
Un même bloc arithmétique est utilisé par N opérateurs ; un pas de calcul est effectué en traitant successivement ces N opérateurs par le bloc arithmétique ; les opérateurs sont traités dans l'ordre où ils sont numérotés.

D'autre part, l'ADD étudié ne dispose que d'une mémoire d'incrément.

Nous noterons $a_k(i)$ l'arrondi au pas i relatif à l'opérateur (k) . Si l'incrément émis par l'opérateur (k) a un poids $2^{p_k} \cdot h$

$$a_k(i) = 2^{p_k} h F(R_i^k)$$

* Supposons, dans un but de simplification que la chaîne constitue une boucle simple



Nous appellerons erreur d'arrondi au pas i sur la valeur Y_i^k , l'erreur provoquée par la propagation des arrondis au même pas i (pour les opérateurs 1 à $k-1$) et au pas précédent $i-1$ (pour les opérateurs $k, k+1, \dots, n$)

Le calcul commence à l'opérateur n° 1

Nous noterons cette erreur d'arrondi $\epsilon_a^k(i)$

$$\epsilon_a^k(i) = f \left[a_1(i), a_2(i), \dots, a_{k-1}(i); a_k(i-1), \dots, a_n(i-1) \right]$$

Nous noterons ϵ_o^k l'erreur initiale ($\epsilon_o^k = 2^{q_k} Y_o^k - y_o^k$)

et $\epsilon_r^k(i)$ l'erreur de retard au pas i .

* On peut prévoir une propagation de ces erreurs initiales, d'arrondi et de retard en cours de calcul.

Nous noterons $\epsilon_{p_o}^k(i)$, $\epsilon_{p_a}^k(i)$ et $\epsilon_{p_r}^k(i)$ ces erreurs

de propagation.

d- But de l'étude.

L'étude que nous allons exposer est essentiellement une étude de la propagation des arrondis en cours de calcul.

Nous avons distingué

- L'erreur d'arrondi au pas i qui a pour source les arrondis aux pas i ou $i - 1$ (suivant le numéro de l'opérateur)
- L'erreur due à la propagation des erreurs d'arrondis au cours du calcul des pas $i-2, i-3 \dots 0$.

Le but de cette étude est la réduction des erreurs d'arrondi et des erreurs de propagation.

Nous avons étudié quelques problèmes simples de résolution d'équations différentielles ; le choix de ces équations différentielles a été fait de manière à constituer des circuits où sont utilisés les opérateurs essentiels de l'analyseur différentiel digital.

- | | |
|--|---|
| Chap. I et II : étude de $y' = \pm y$ | Intégration par rapport à la variable indépendante |
| Chap. III et IV : étude de $y' = \pm ay$ | Intégration par rapport à la variable indépendante et multiplication par une constante. |
| Chap. V : étude de $y'' = -y$ | Deux intégrations en chaîne par rapport à la variable indépendante. |
| Chap. VI : résolution de $ax^2 + bx + c = 0$ | |

Intégration par rapport à la variable indépendante, intégration par rapport à une variable quelconque, 2 multiplications par une constante.

Remarque: Dans les chap. V et VI nous avons fait une étude rapide de l'erreur de retard.

Rappelons que l'analyseur différentiel digital étudié a les particularités suivantes :

- il est de type séquentiel
- il a une seule mémoire d'incrément
- la numération est binaire
- l'émission d'incrément est faite de la manière suivante :

$$E(R_i) = + 1 \quad \text{si} \quad R_i \geq + 1$$

$$E(R_i) = 0 \quad \text{si} \quad + 1 > R_i > - 1$$

$$E(R_i) = - 1 \quad \text{si} \quad R_i < - 1$$

CHAPITRE I

ETUDE DE $y' = -y$

I. A.1 Expression théorique de l'erreur de principe dans la résolution de $y' = -y$.

Nous avons vu dans le chapitre précédent que nous pouvions écrire Y_i sous la forme :

$$Y_i = Y_{i-1} - h Y_{i-1} + h \lambda_{i-1} \quad (1)$$

avec $\lambda_{i-1} = F(R_{i-1}) - F(R_{i-2})$ et $\lambda_0 = F(R_0)$

L'erreur de principe $\epsilon_i = 2^q$. $Y_i - y_i$ au pas i peut être obtenue en comparant la relation

$$y_i = y_{i-1} - h y_{i-1}$$

avec l'expression (1) de Y_i

Ainsi

$$\epsilon_i = \epsilon_{i-1} (1-h) + h 2^q \lambda_{i-1}$$

$$\text{soit } \epsilon_i = (1-h)^i \epsilon_0 + h 2^q \left[\lambda_{i-1} + (1-h) \lambda_{i-2} + \dots + (1-h)^{i-1} \lambda_0 \right]$$

Comme $\lambda_j = F(R_j) - F(R_{j-1})$

$$\epsilon_i = (1-h)^i \epsilon_0 + h 2^q F(R_{i-1}) - h^2 2^q \left[F(R_{i-2}) + (1-h)F(R_{i-3}) + \dots + (1-h)^{i-2} F(R_0) \right]$$

L'addition de Y à R a lieu à chaque pas d'intégration (intégration par rapport à la variable indépendante).

On peut admettre que la moyenne des valeurs $F(R_j)$ dans un intervalle $[0, i]$ est égale à $+\frac{1}{2}$ (même avec i assez petit)

Dans ces conditions

$$\epsilon_i = (1-h)^i \epsilon_0 + h 2^q F(R_{i-1}) - h 2^{q-1} \left[1 - (1-h)^{i-1} \right]$$

L'arrondi au pas $i-1$ est $a(i-1) = h 2^q F(R_{i-1})$

Le circuit ne comprenant qu'un seul intégrateur, l'erreur d'arrondi au pas i est

$$\epsilon_a (i) = a(i-1) = h 2^q F(R_{i-1})$$

en utilisant les notations définies dans l'introduction

$$\epsilon_i = \epsilon_{p_0} (i) + \epsilon_a (i) + \epsilon_{p_a} (i)$$

En passant à la variable $x_i = x_0 + ih$ et en posant

$$-\alpha = \frac{L(1-h)}{h}$$

$$\epsilon_{p_0} (x_i) + \epsilon_{p_a} (x_i) = \epsilon_0 e^{-\alpha (x_i - x_0)} - h 2^{q-1} \left[1 - e^{-\alpha (x_{i-1} - x_0)} \right]$$

La courbe ayant pour équation

$$p(x_i) = \epsilon_0 e^{-\alpha (x_i - x_0)} - h 2^{q-1} \left[1 - e^{-\alpha (x_{i-1} - x_0)} \right]$$

peut être appelée courbe de propagation

ainsi
$$\epsilon (x_i) = \epsilon_a (x_i) + p (x_i)$$

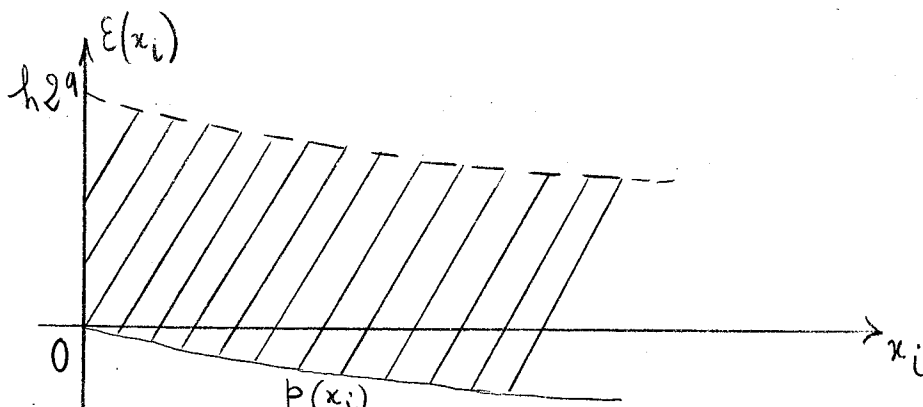
Nous savons que :

$$0 \leq \epsilon_a (x_i) < h 2^q$$

Les limites de variations de $\epsilon (x_i)$ sont donc parfaitement définies

$$p (x_i) \leq \epsilon (x_i) < p (x_i) + h 2^q$$

Graphiquement, la courbe représentative de l'erreur de principe $\epsilon (x_i)$ appartient à une bande du plan de largeur $h 2^q$ et ayant la courbe de propagation pour plancher.



Nous remarquerons que, quelque soit ϵ_0 , $p(x_i)$ tend vers $-h 2^{q-1}$; la courbe de propagation ne dérive donc pas.

I. A 2 Erreur moyenne dans un intervalle $[0, i]$

Le point de la courbe d'erreur d'abscisse x_i a pour ordonnée

$$\bar{\epsilon}(x_0, x_i) \neq \frac{1 - e^{-\alpha(x_i - x_0)}}{\alpha(x_i - x_0)} \quad [\epsilon_0 + h 2^{q-1}]$$

I. A 3 Vérification numérique

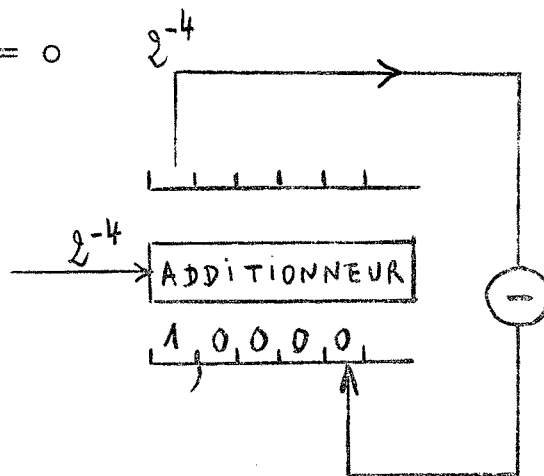
Une première étude de $y' = -y$ a été faite avec un pas $h = 2^{-4}$ et des valeurs initiales

$$\begin{cases} y_0 = 1 ; & \epsilon_0 = 0 \\ x_0 = 0 \end{cases}$$

nous avons pris $Y_0 = 1,0000$

(dans un ADD, $-1 \leq Y_j < 1$; ici

nous avons admis que Y_j pouvait être égal à 1 de manière à éviter parfois des problèmes de cadrage)



Ainsi $q = 0$, $\alpha = \frac{-L(1-h)}{h} = 1,031$

et, théoriquement,

$$\epsilon(x_i) = -2^{-5} (1 - e^{-1,031 x_i}) + \epsilon_a(x_i) \text{ avec } 0 \leq \epsilon_a(x_i) < 2^{-4}$$

$$\frac{\epsilon(x_0, x_i)}{x_i} = \frac{2^{-5}}{x_i} (1 - e^{-1,031 x_i})$$

Graphiques

Sur un premier graphique G I A₁, est représentée la courbe intégrale obtenue en utilisant l'ADD.

La deuxième courbe intégrale (cas ②) que l'on peut voir sur ce graphique correspond à une étude, que nous exposerons plus loin, sur la réduction de l'erreur de principe.

Sur le graphique G.I.A₂ est tracée la courbe d'erreur réelle.

La courbe d'erreur varie dans la zone définie théoriquement.

Nous avons fait un essai en prenant $\epsilon_0 = -2^{-4}$

(graphique G I A₃) la courbe de propagation a alors pour équation

$$p(x_i) = -2^{-5} (1 + e^{-1,031 x_i})$$

Une deuxième vérification a été faite avec un pas $h = 2^{-6}$

(graphique G.I.A₄ pour la courbe d'erreur)

I.A.4 Erreur de principe et erreur de méthode

Soit (y_i, x_i) un point exact de la courbe intégrale étudiée. Supposons que les valeurs initiales sont $x_0 = 0, y_0 = 1$; alors $y_i = e^{-x_i}$ Soit y_i l'approximation de y_i donnée par la méthode de la tangente.

$$y_i = y_{i-1} - h y_{i-1}$$

$$\text{alors } y_i = (1-h)^i = e^{-\alpha x_i} \quad \text{avec } -\alpha = \frac{L(1-h)}{h}$$

Nous avons défini au chapitre précédent ^{l'erreur} de méthode $\sigma(x_i)$

$$\begin{aligned} \sigma(x_i) &= y_i - \hat{y}_i \\ &= e^{-\alpha x_i} - e^{-x_i} \sim -\frac{h^2 x_i}{3!} < 0 \text{ pour} \end{aligned}$$

$$x_i > 0$$

d'autre part $\epsilon(x_i) = h 2^{q-1} \left[\pm 1 + e^{-\alpha x_i} \right]$ en supposant que $\epsilon_0 = 0$

Sauf dans le cas où $q \ll 0$, $|\epsilon(x_i)| \gg |\sigma(x_i)|$

I - B I Réduction de l'erreur de principe.

Nous avons vu dans l'introduction que si, à chaque pas, l'intégrateur faisait le calcul

$$Y_i = Y_{i-1} - h R_{i-1}$$

au lieu de

$$Y_i = Y_{i-1} - h E (R_{i-1})$$

et que $\epsilon_0 = 0$

il n'y aurait pas d'erreur de principe dans la résolution de $y' = -y$

Nous avons appelé arrondi au pas $i - 1$ la valeur $a(i-1) = h 2^q F (R_{i-1})$

Dans la résolution de $y' = -y$ le registre R de l'intégrateur utilisé contient toujours des valeurs R_i positives

$$0 \leq R_i < 2$$

Supposons qu'une valeur $+\frac{1}{2}$ soit mise initialement dans le registre R et notons :

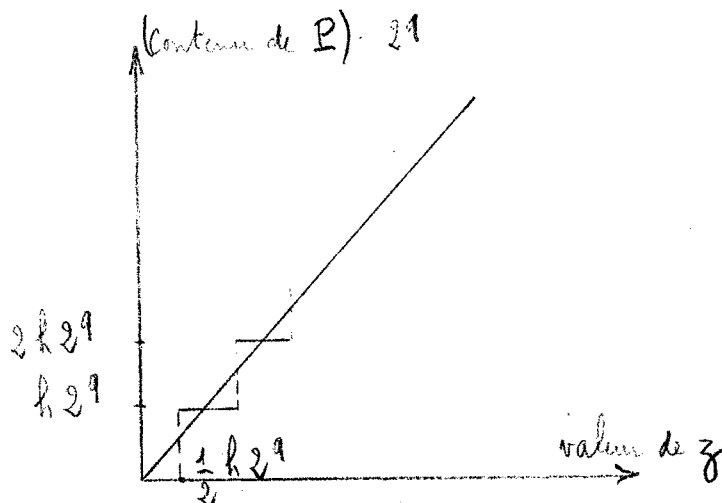
R'_i la valeur du contenu de R au pas i

Y'_i la valeur du contenu de Y au pas i

L'arrondi au pas i vaut dans ce cas

$$a(i) = h 2^q \left[F (R'_i) - \frac{1}{2} \right]$$

Cet arrondi peut être représenté graphiquement comme suit



Au pas i ,

$$z_i = h 2^q \sum_0^i E(R_K) + \underbrace{h 2^q F(R_i) - \frac{1}{2} h 2^q}_{\text{arrondi } a(i)}$$

l'arrondi varie donc entre les valeurs $-h 2^{q-1}$ et $+h 2^{q-1}$

Nous avons vu que $\epsilon_a(i) = a(i-1)$, Les arrondis prenant des valeurs positives ou négatives, on peut prévoir que les erreurs d'arrondi aux pas $0, 1, 2, \dots, i-1$ se compenseront de telle sorte qu'il n'y aura plus d'erreur de propagation $\epsilon_{p_a}(i)$ au pas i .

Nous avons vu que l'erreur de principe ϵ'_i au pas i pouvait s'écrire sous la forme

$$\epsilon'_i = (1-h)^i \epsilon_0 + h 2^q \left[\lambda'_{i-1} + (1-h) \lambda'_{i-2} + \dots + (1-h)^{i-1} \lambda'_0 \right]$$

$+ \frac{1}{2}$ étant mis initialement dans le registre R , λ'_0 vaut

$$\lambda'_0 = F(R'_0) - \frac{1}{2}$$

Alors, en supposant que $F[\overline{R'_j}]_{[0, i-1]} = \frac{1}{2}$

$$\epsilon'_i = (1-h)^i \epsilon_0 + h 2^q F(R'_{i-1}) - h 2^{q-1} \left[1 - (1-h)^{i-1} \right] - h 2^{q-1} (1-h)^{i-1}$$

$$\text{soit : } \epsilon'_i = (1-h)^i \epsilon_0 + h 2^q \left[F(R'_{i-1}) - \frac{1}{2} \right]$$

soit encore

$$\boxed{\epsilon'_i = \epsilon_{p_0}(i) + \epsilon_a(i)}$$

Remarquons que ce calcul n'est qu'une vérification de l'exactitude de l'expression de l'erreur propagée $\epsilon_{p_a}(i)$

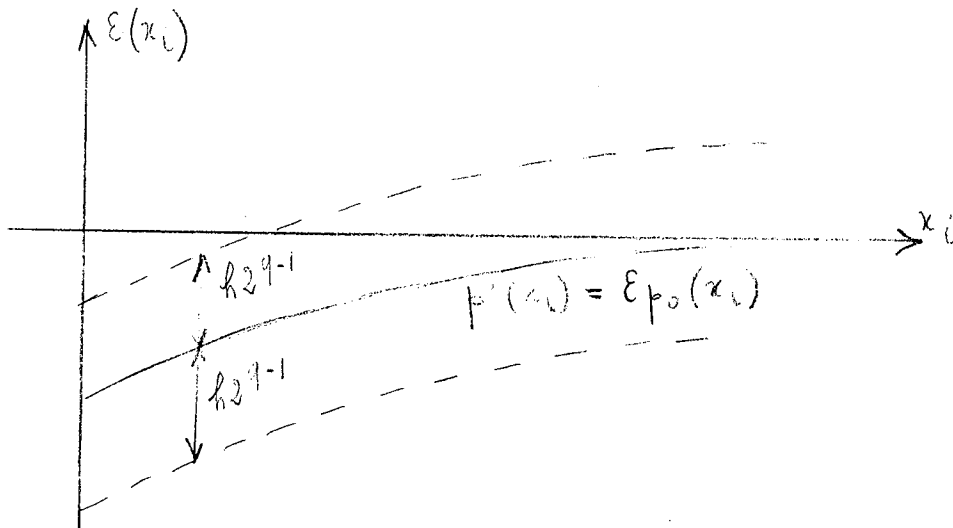
Dans le paragraphe I.A1, l'expression de $\epsilon_{p_a}(i)$ a été donné en supposant que la valeur moyenne dans $[0, i]$ de l'erreur d'ar-

rondi était $\frac{1}{2} \cdot h 2^q$. ; dans le cas présent, on suppose que cette valeur moyenne est nulle, il est donc normal que $\epsilon'_{Pa}(i) = 0$ quel que soit i .

Graphiquement la courbe représentative de l'erreur de principe $\epsilon(x_i)$ appartient à une bande du plan ayant pour largeur $h 2^q$ et dont la tendance générale est celle de la courbe de propagation de l'erreur initiale ϵ_0

Remarque : Quel que soit ϵ_0 , $P'(x_i)$ tend vers zéro.

Exemple : Pour $\epsilon_0 < 0$, $x_0 = 0$



$$P'(x_i) - h 2^{q-1} < \epsilon'(x_i) < P'(x_i) + h 2^{q-1}$$

On voit immédiatement que dans le cas où ϵ_0 est nul, l'erreur de principe au pas i est égale à l'erreur d'arrondi à ce même pas i .

Les bornes de variation de l'erreur de principe sont alors parfaitement définies $(-h 2^{q-1}, +h 2^{q-1})$ et minimisées en valeur absolue.

Avant de vérifier expérimentalement ces résultats, je voudrais faire remarquer que la diminution de la valeur absolue de l'erreur d'arrondi (que nous appellerons cas (2)) n'est pas toujours la meilleure méthode pour minimiser l'erreur de principe.

Supposons en effet que $\epsilon_0 = -h 2^{q-1}$ (la position extrême droite de Y_0 a un poids $h 2^{-q}$)

Dans le cas ① où $0 \leq \epsilon_a(j) < h 2^q$

$$p(x_i) = -h 2^{q-1} e^{-\alpha(x_i-x_0)} - h 2^{q-1} (1 - e^{-\alpha(x_{i-1}-x_0)})$$

$$p(x_i) \neq -h 2^{q-1}$$

$$\text{et } -h 2^{q-1} < \epsilon(x_i) < h 2^{q-1}$$

Dans le cas ② où $-h 2^{q-1} < \epsilon'_a(j) < +h 2^{q-1}$

$$p'(x_i) = -h 2^{q-1} e^{-\alpha(x_i-x_0)}$$

$$\text{et } -h 2^{q-1} [1 + e^{-\alpha(x_i-x_0)}] < \epsilon'(x_i) < h 2^{q-1} [(1 - e^{-\alpha(x_i-x_0)})]$$

dans ce cas, évidemment assez particulier, il est préférable de laisser varier l'erreur d'arrondi entre 0 et $h 2^q$.

I-B-2 Vérification numérique

Deux expériences ont été faites avec les pas $h = 2^{-4}$ (G.I.B1) et $h = 2^{-6}$ (G.I.B2) et avec les valeurs initiales

$$x_0 = 0, y_0 = 1, q = 0, \epsilon_0 = 0$$

L'erreur moyenne a une valeur théorique nulle [G.IA₅ (cas 2)]. En fait la courbe d'erreur moyenne est située sous l'axe des x_i .

Ce phénomène correspond à une mauvaise estimation de $F(R_j)$ (dans le paragraphe I-C nous verrons qu'il s'agit d'une surestimation).

En mettant une valeur initiale $K = 0,4375$ dans le registre R, (cas ③ sur G.I. A5) la courbe d'erreur moyenne obtenue oscille sur l'axe des x .

I - C Expression générale de l'erreur de principe

Supposons qu'une valeur k , ($-1 < k < 1$) soit mise initialement dans le registre R de l'intégrateur.

Au pas i ,

valeur de l'intégrale $z = h 2^q$ (contenu du registre Z) - $h 2^q \cdot k$

$= h 2^q$ (contenu de P + contenu de R) - $h 2^q \cdot k$

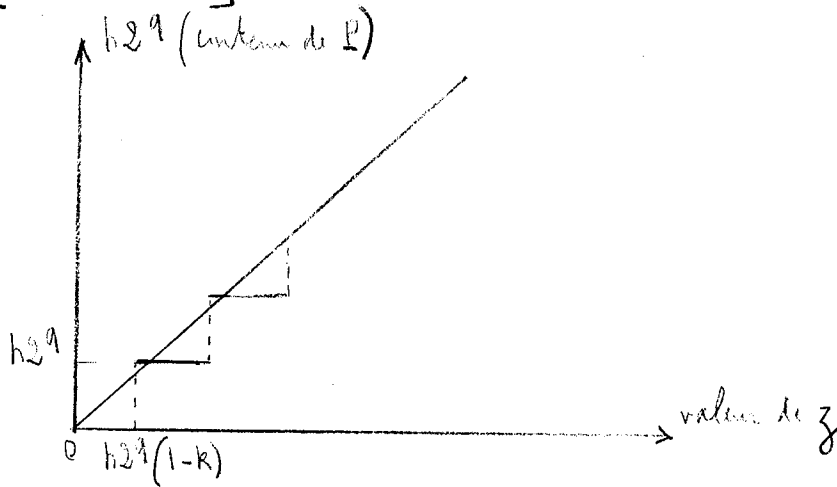
$$= h 2^q \underbrace{\sum_0^i E(R_k)}_{\text{contenu de } p} + h 2^q \underbrace{(F(R_i) - k)}_{a(i)}$$

contenu de p

$a(i)$

l'arrondi sur la valeur de l'intégrale au pas i est donc

$$a(i) = h 2^q [F(R_i) - k]$$



Quel que soit la valeur k , on peut écrire :

$$\epsilon_i = (1-h)^i \epsilon_0 + h 2^q \left[\lambda_{i-1} + (1-h) \lambda_{i-2} + \dots + (1-h)^{i-1} \lambda_0 \right]$$

avec $\lambda_0 = F(R_0) - k$

$$\epsilon_i = (1-h)^i \epsilon_0 + h 2^q F(R_{i-1}) - h 2^q \overline{F(R_j)}_{[0, i-1]} (1-(1-h)^{i-1}) - h 2^q \cdot k (1-h)^{i-1}$$

$$\epsilon_i = (1-h)^i \epsilon_0 + h 2^q [F(R_{i-1}) - k] - h 2^q [1 - (1-h)^{i-1}] \left[\overline{F(R_j)}_{[0, i-1]} - k \right]$$

soit encore $\epsilon_i = \epsilon_{p_0}(i) + \epsilon_a(i) + \epsilon_{pa}(i)$ avec $-h 2^q k \leq \epsilon_a(i) < h 2^q (1-k)$

Il suffit de remplacer K par zéro ou $\frac{1}{2}$, dans l'expression écrite plus haut et de faire l'hypothèse $\overline{F(R_j)} = \frac{1}{2}$ pour retrouver les expressions déjà données de ϵ_i .

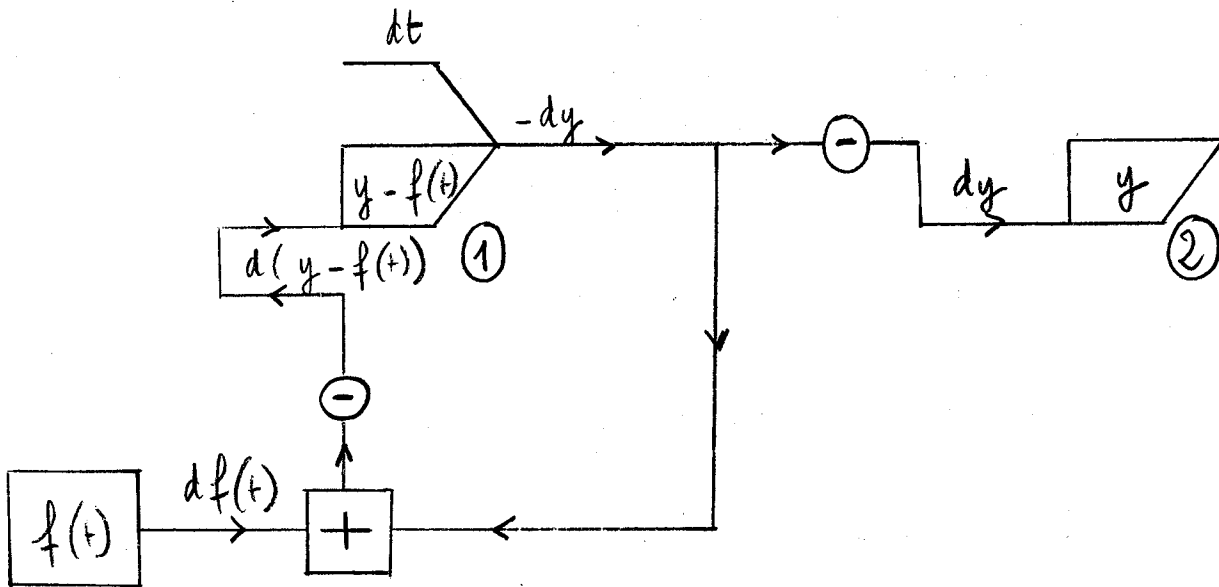
Remarque : On montrerait de même que :

$$\overline{\epsilon}_{(x_0, x_i)} = \frac{1 - e^{-\alpha (x_i - x_0)}}{\alpha (x_i - x_0)} \left[\epsilon_0 + h 2^q \left[\overline{F(R_j)} - k \right] \right]$$

Dans les études que nous ferons par la suite nous supposerons toujours qu'une valeur K_j a été mise initialement dans le registre R^j d'un opérateur (j).

Remarque : Résolution d'une équation de la forme $y' = -y + f(t)$

Le schéma représentatif est le suivant :



Les incréments représentatifs de $-y \, dt$ et de $f(t)dt$ sont additionnés avant d'alimenter le registre Y^1 . Cette addition se fait dans un additionneur représenté par la figure $\boxed{+}$. Les incréments doivent avoir même poids.

Les incréments émis par l'intégrateur (1) sont accumulés dans un registre Y^2 . La solution $y(t)$ est représentée à chaque pas par le contenu Y^2_i de ce registre. Si q_1 et q_2 sont les facteurs de cadrage relatifs aux opérateurs (1) et (2), c'est-à-dire, qu'au début du calcul les registres Y^1 et Y^2 contiennent respectivement $Y^1_0 = 2^{-q_1} (y_0 - f(t_0))$ et $Y^2_0 = 2^{-q_2} y_0$

les relations fondamentales sont :

$$\text{pour l'intégrateur (1), } Y^1_i = Y^1_{i-1} - h \left[E(R^1_{i-1}) + E(f(t_{i-1})) \right]$$

$$\text{pour le registre } Y^2, \quad Y^2_i = Y^2_{i-1} - h 2^{q_1 - q_2} E(R^1_{i-1})$$

Pour que cette dernière relation soit vraie, il faut considérer que le registre Y^2 contient à la fin du pas i , non pas la valeur $2^{-q_2} y_i$ mais la valeur $2^{-q_2} y_{i+1}$. Au pas 0, en effet, l'intégrateur peut émettre un incrément $E(R^1_0)$ et modifier le registre Y^2 .

Ces relations fondamentales correspondent bien au système :

$$\begin{cases} y'_i = y'_{i-1} + h y''_{i-1} \\ y_i = y_{i-1} + h y'_{i-1} \end{cases}$$

c'est-à-dire en posant $-y'_i = y^1_i$ et $y_i = y^2_i$, au système :

$$\begin{cases} y^1_i = y^1_{i-1} - h \left(y^1_{i-1} + f'(t_{i-1}) \right) \\ y^2_i = y^2_{i-1} - h y^1_{i-1} \end{cases}$$

En effet, les relations fondamentales peuvent encore s'écrire :

$$Y^1_i = Y^1_{i-1} - h \left[Y^1_{i-1} + E(f(t_{i-1})) \right] + h \lambda^1_{i-1}$$

$$Y^2_i = Y^2_{i-1} - h 2^{q_1 - q_2} Y^1_{i-1} + h 2^{q_1 - q_2} \lambda^1_{i-1}$$

Si nous notons les erreurs de principe $\epsilon^1_i = 2^{q_1} \cdot Y^1_i - y^1_i$

et $\epsilon^2_i = 2^{q_2} \cdot Y^2_i - y^2_i$

et si nous supposons que $f'(t)$ est une fonction discrète représentable parfaitement par un incrément (c'est-à-dire $2^{q_1} E(f(t_{i-1})) = f'(t_{i-1}))$)

le système d'erreur s'écrit :

$$\begin{cases} \epsilon^1_i = \epsilon^1_{i-1} - h \epsilon^1_{i-1} + h 2^{q_1} \lambda^1_{i-1} \\ \epsilon^2_i = \epsilon^2_{i-1} - h \epsilon^1_{i-1} + h 2^{q_1} \lambda^1_{i-1} \end{cases}$$

ainsi $\epsilon^1_j = \epsilon^2_j \quad \forall j$ (nous supposons que $\epsilon^1_0 = \epsilon^2_0$)

le système (1) devient

$$\epsilon^1_i = \epsilon^1_{i-1} (1-h) + h 2^{q_1} \lambda_{i-1} = \epsilon^2_i$$

qui est la relation qui détermine l'erreur de principe dans la résolution de $y' = -y$. En conclusion l'erreur de principe a les mêmes particularités dans la résolution de $y' = -y$ et $y' = -y + f(t)$ avec les hypothèses restrictives suivantes :

- $f(t)$ est supposée connue sans erreur d'arrondi
- on considère que le registre Y^2 est en avance d'un pas sur l'opérateur (1)

c'est-à-dire que, à la fin du cycle de calcul i , le registre Y^2 contient non pas la valeur $2^{q_2} y_i$ mais la valeur $2^{-q_2} y_{i+1}$

I - D. RESOLUTION DE $y' = -y$ PAR LA METHODE DES TRAPEZES

La méthode d'intégration des trapèzes est couramment utilisée. Les valeurs successives de y_i sont obtenues à partir de la relation

$$y_i = y_{i-1} - h \left(y_{i-1} + \frac{1}{2} (y_{i-1} - y_{i-2}) \right) \quad (1)$$

A cette relation correspondent les relations fondamentales suivantes :

$$\begin{cases} Y_i = Y_{i-1} - h E(R_{i-1}) \\ R_{i-1} = F(R_{i-2}) + Y_{i-1} - \frac{1}{2} h E(R_{i-2}) \end{cases}$$

Comme

$$Y_{i-1} = Y_{i-2} - h E(R_{i-2})$$

on a :

$$R_{i-1} = F(R_{i-2}) + Y_{i-1} + \frac{1}{2} (Y_{i-1} - Y_{i-2})$$

Ainsi la relation (1) est bien représentée par :

$$Y_i = Y_{i-1} - h \left[Y_{i-1} + \frac{1}{2} (Y_{i-1} - Y_{i-2}) \right] + h \lambda_{i-1}$$

L'erreur de principe vaut

$$\epsilon_i = \epsilon_{i-1} - h \left(\epsilon_{i-1} + \frac{1}{2} (\epsilon_{i-1} - \epsilon_{i-2}) \right) + h 2^q \lambda_{i-1}$$

soit :

$$\epsilon_i = \epsilon_{i-1} \left(1 - \frac{3}{2} h \right) + \frac{1}{2} h \epsilon_{i-2} + h 2^q \lambda_{i-1}$$

ou encore :

$$\epsilon_i = \alpha \delta_{i-1} \epsilon_0 + h 2^q \sum_{j=1}^i \delta_{j-1} \lambda_{i-j}$$

où δ_j est donné par la relation de récurrence

$$\delta_j = \alpha \delta_{j-1} + \beta \delta_{j-2}$$

avec

$$\delta_0 = 1, \delta_1 = \alpha$$

et

$$\alpha = 1 - \frac{3}{2} h, \quad \beta = \frac{h}{2}$$

Ainsi :

$$\epsilon_i = \alpha \delta_{i-1} \epsilon_0 + h 2^q F(R_{i-1}) + h 2^q \left[F(R_{i-2})(\delta_{i-1} - \delta_0) + \dots + F(R_0)(\delta_{i-1} - \delta_{i-2}) - k \delta_{i-1} \right]$$

soit encore :

$$\epsilon_i = \alpha \delta_{i-1} \epsilon_0 + h 2^q \left[F(R_{i-1}) - k \right] + h 2^q \left[\overline{F(R_j)}_{(0, i-2)} - k \right] (\delta_{i-1} - \delta_0)$$

Le terme $\delta_{i-1} - \delta_0$ est une fonction $\theta(h)$

Ainsi :

$$\epsilon_i = \underbrace{\alpha \delta_{i-1} \epsilon_0}_{\epsilon_{po}(i)} + \underbrace{h 2^q (F(R_{i-1}) - k)}_{\epsilon_a(i)} + \underbrace{h 2^q (\overline{F(R_j)}_{(0, i-2)} - k)}_{\epsilon_{pa}(i)} \theta(h)$$

On remarque que :

- l'erreur d'arrondi a la même expression que lorsque la méthode des rectangles est utilisée.

- l'erreur $\epsilon_{pa}(i)$ peut être éliminée en mettant une valeur k convenable dans le registre R avant tout calcul.

Sur le graphique G.I.D, on peut constater qu'en mettant une valeur initiale $k = \frac{1}{2}$ dans R et en intégrant par la méthode des trapèzes, les bornes de variation de ϵ_i sont $h 2^{q-1}$ et $-h 2^{q-1}$.

CHAPITRE II

ETUDE DE $y' = + y$

II-1 Etude théorique

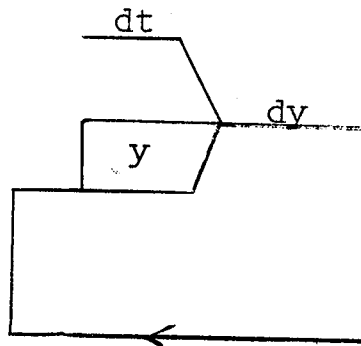
Nous avons vu dans le chapitre précédent que la courbe de propagation des erreurs initiale et d'arrondi avait une équation qui s'exprimait en fonction (ou presque) de la solution de $y' = - y$. Cette solution ne dérivait pas et la courbe de propagation non plus.

Nous allons donc étudier l'équation différentielle $y' = + y$ dont les solutions dérivent.

La relation de récurrence à représenter est :

$$y_i = y_{i-1} + h y_{i-1}$$

et le schéma de principe est le suivant



Les relations fondamentales entre les valeurs contenues dans les registres sont

$$\begin{cases} Y_i = Y_{i-1} + h E(R_{i-1}) & E(R_{i-1}) = 0 \text{ ou } 1 \\ R_i = F(R_{i-1}) + Y_i \end{cases}$$

donc $Y_i = Y_{i-1} + h Y_{i-1} - h \lambda_{i-1}$

L'erreur de principe au pas i peut être évalué de la même manière qu'au chapitre précédent.

Cette erreur ϵ_i vaut :

$$\epsilon_i = (1+h)^i \epsilon_0 - h2^q [F(R_{i-1}) - k] + h2^q (1-(1+h)^{i-1}) \left[\overline{F(R_j)}_{(0,i-1)} - k \right]$$

L'arrondi sur l'incrément émis par l'intégrateur au pas $i-1$ est :

$$a(i-1) = h2^q [F(R_{i-1}) - k]$$

donc :

$$\epsilon_a(i) = -a(i-1)$$

La relation peut s'écrire

$$\epsilon_i = \epsilon_{P_0}(i) + \epsilon_a(i) + \epsilon_{P_a}(i)$$

En passant à la variable $x_i = x_0 + ih$ et en posant

$$\alpha = \frac{L(1+h)}{h}$$

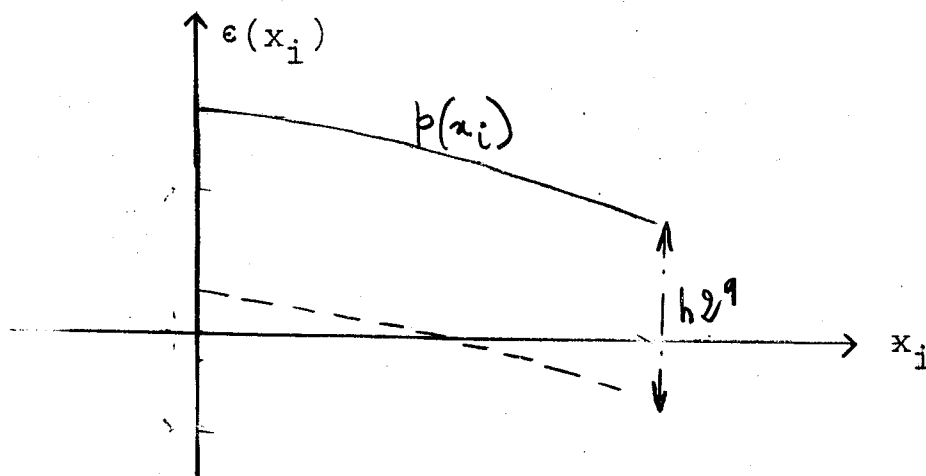
la courbe de propagation a pour équation

$$p(x_i) = \epsilon_0 e^{\alpha(x_i - x_0)} + h2^q \left[1 - e^{\alpha(x_{i-1} - x_0)} \right] \left[\overline{F(R_j)}_{(0,i-1)} - k \right]$$

Cas 1

Si $k = 0$ et $\overline{F(R_j)} = \frac{1}{2}$ $p(x_i) = \epsilon_0 e^{\alpha(x_i - x_0)} + h2^{q-1} (1 - e^{\alpha(x_{i-1} - x_0)})$

et $-h2^q < \epsilon_a(x_i) \leq 0$



On remarque que la courbe de propagation dérive. Evidemment à partir de l'expression de $p(x_i)$, on peut déterminer la valeur maximum de l'erreur de principe puisque le calcul est limité par la longueur du registre Y.

Cependant, on ne pourra faire un calcul sain que si ϵ_{Pa} est éliminée

Cas 2

Nous avons fait l'hypothèse $\overline{F(R_j)} [x_0, x_{i-1}] = \frac{1}{2} \sqrt{x_{i-1}}$

Si cette hypothèse est correcte, il suffit de prendre $k = \frac{1}{2}$ pour éliminer le terme $\epsilon_{Pa}(x_i)$

dans ce cas :

$$\epsilon_{(x_i)} = \epsilon_0 e^{-\alpha(x_i - x_0)} - h 2^q \left[F(R_{i-1}) - \frac{1}{2} \right]$$

Dans le cas où $\epsilon_0 = 0$, l'erreur de principe est égale à l'erreur d'arrondi. Ce résultat est intéressant car l'erreur de principe est considérablement atténuée.

II-2 Vérification numérique

Nous avons résolu $y' = +y$ en prenant $y_0 = 1, x_0 = 0, \epsilon_0 = 0 ; q = 3, h = 2^{-7}$

Cas 1 :

$$k = 0$$

$$\epsilon_{(x_i)} = 2^{-5} (1 - e^{0,997 x_{i-1}}) - \epsilon_a(x_i)$$

$$0 \leq \epsilon_a(x_i) < 2^{-4}$$

graphique G II 1

Cas 2 :

$$k = \frac{1}{2}$$

$$\epsilon(x_i) = \epsilon_a(x_i) \left[-2^{-5} < \epsilon_a(x_i) < 2^{-5} \right]$$

Graphique G II 2

Nous remarquerons que la théorie n'est plus parfaitement vérifiée lorsque x_i croît. Ceci est normal car dans l'expression de $\epsilon(x_i)$ intervient le terme

$$h 2^q (1 - e^{-\alpha(x_{i-1} - x_0)}) (\overline{F(R_j)} - k)$$

lorsque $k = \frac{1}{2}$ ce terme disparaît si $\overline{F(R_j)} = \frac{1}{2}$ très exactement.

$$\text{En fait : } \overline{F(R_j)} - \frac{1}{2} \neq 0$$

et $h 2^q (1 - e^{-\alpha(x_{i-1} - x_0)}) (\overline{F(R_j)} - \frac{1}{2})$ a une valeur absolue qui croît avec x .

A la vue de la courbe réelle, on peut estimer que

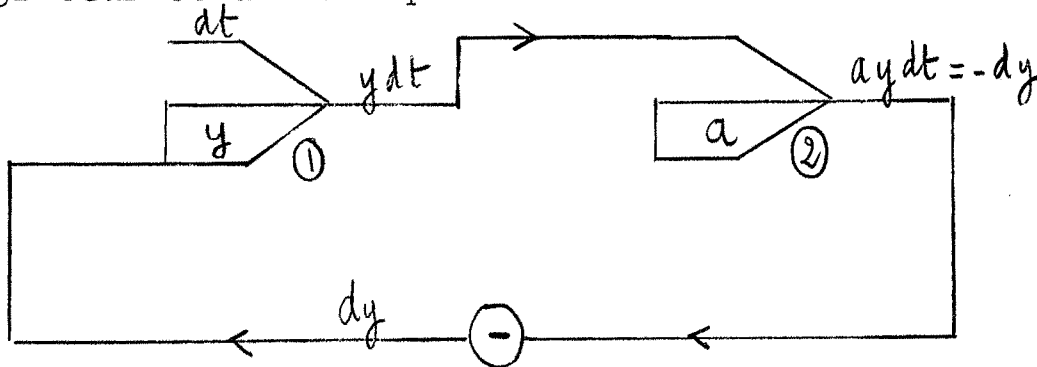
$$\overline{F(R_j)} < \frac{1}{2}$$

- Les courbes d'erreurs moyennes sont tracées sur le graphique G II 3.

CHAPITRE III

ETUDE DE $y' = -ay$ ($a > 0$)

Le circuit représentatif de cette équation comprend un intégrateur et un multiplieur.



Rappelons brièvement le principe du multiplieur.

L'addition ou la soustraction du registre Y au registre R est commandée par l'incrément émis par un autre opérateur.

L'incrément émis au pas i est donc une représentation de

$$\int_{x_{i-1}}^{x_i} K dx \quad \text{si } K, \text{ constant, est contenu dans Y et si l'incrément émis est une représentation de la différentielle } dx$$

Dans le cas où le registre Y contient une représentation d'une fonction y_i , on calcule une intégrale au sens de Stieljes.

- Le circuit représente la relation

$$y_i^1 = y_{i-1}^1 - ah y_{i-1}^1 \quad (\text{en posant } y_j = y_j^1)$$

III-A-1 Relations fondamentales.

L'incrément émis par l'intégrateur (1) a un poids $2^{q_1} h$

Cet incrément commande le fonctionnement du multiplieur 2 ; l'incrément émis par ce multiplieur a donc un poids $(2^{q_1} h) 2^{q_2}$

Les relations fondamentales peuvent s'écrire

pour l'intégrateur (1)
$$\begin{cases} Y_i^1 = Y_{i-1}^1 - 2^{-q_1} [2^{q_1 + q_2} h E(R_{i-1}^2)] \\ R_i^1 = F(R_{i-1}^1) + Y_i^1 \end{cases}$$

pour le multiplieur (2)
$$\begin{cases} Y_i^2 = Y_0^2 = 2^{-q_2} \cdot a \\ R_i^2 = F(R_{i-1}^2) + E(R_i^1) Y_{i-1}^2 \end{cases}$$

En faisant appel aux variables

$$\lambda_i^1 = F(R_i^1) - F(R_{i-1}^1) \quad \text{et} \quad \lambda_i^2 = F(R_i^2) - F(R_{i-1}^2)$$

Y_i^1 peut s'écrire

$$Y_i^1 = Y_{i-1}^1 - h 2^{q_2} Y_0^2 Y_{i-1}^1 + h [2^{q_2} Y_0^2 \lambda_{i-1}^1 + 2^{q_2} \lambda_{i-1}^2]$$

III-A-2 Expressions théoriques de l'erreur de principe sur la valeur contenue dans le registre Y^1

Si les erreurs initiales sont ϵ_0^1 et ϵ_0^2 ,

l'erreur de principe $\epsilon_i^1 = 2^{q_1} Y_i^1 - y_i^1$ vaut

$$\epsilon_i^1 = \epsilon_{i-1}^1 - ah \epsilon_{i-1}^1 + 2^{q_1} h \epsilon_0^2 (\lambda_{i-1}^1 - Y_{i-1}^1) + 2^{q_1} ah \lambda_{i-1}^1 + 2^{q_1 + q_2} h \lambda_{i-1}^2$$

on remarquera que l'erreur ϵ_o^2 est non seulement propagée mais fausse le calcul à chaque pas d'intégration.

L'erreur de principe peut encore s'écrire

$$\begin{aligned} \epsilon_i^1 &= \epsilon_o^2 \cdot 2^{q_1} h \left[\lambda_{i-1}^1 - Y_{i-1}^1 \right] + \epsilon_o^1 (1-ah)^i + \epsilon_o^2 2^{q_1} h \sum_0^{i-2} (1-ah)^j (\lambda_{i-j}^1 - Y_{i-j}^1) \\ &+ 2^{q_1} ah F(R_{i-1}^1) + 2^{q_1+q_2} h F(R_{i-1}^2) - 2^{q_1} a^2 h^2 \left[F(R_{i-2}^1) \right. \\ &+ (1-ah) F(R_{i-3}^1) + \dots + (1-ah)^{i-1} F(R_o^1) \left. \right] \\ &- 2^{q_1+q_2} ah^2 \left[F(R_{i-2}^2) + (1-ah) F(R_{i-3}^2) + \dots + (1-ah)^{i-2} F(R_o^2) \right] \\ &+ 2^{q_1} h (1-ah)^{i-1} \left[a k_1 + 2^{q_2} k_2 \right] \end{aligned}$$

soit encore

$$\begin{aligned} \epsilon_i^1 &= \epsilon_o^1 2^{q_1} h \left[\lambda_{i-1}^1 - Y_{i-1}^1 \right] + \epsilon_o^1 (1-ah)^i \\ &+ \epsilon_o^2 2^{q_1} h \sum_0^{i-2} (1-ah)^j (\lambda_{i-j}^1 - Y_{i-j}^1) \\ &+ 2^{q_1} ah \left[F(R_{i-1}^1) - k_1 \right] + 2^{q_1+q_2} h \left[F(R_{i-1}^2) - k_2 \right] \\ &- 2^{q_1} h \left[1 - (1-ah)^{i-1} \right] a \left[\overline{F(R_j^1)} - k_1 \right] + 2^{q_2} \left[\overline{F(R_j^2)} - k_2 \right] \end{aligned}$$

Simplifions cette expression en faisant les remarques suivantes :

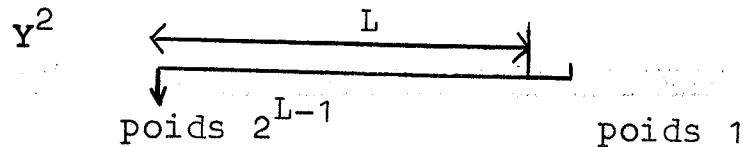
La variable λ_j^1 prend les valeurs $Y_j^1 - 1$ s'il y a un débordement au pas j

et Y_j^1 s'il y a un palier au pas j ;

donc $0 < \lambda_j^1 - Y_j^1 < 1$ quelque soit j

D'autre part, nous savons que le registre d'un multiplieur a une longueur maximum L et que la valeur Y^2_0 peut être mise dans le registre Y^2 avec autant de précision que l'on veut, à concurrence du remplissage du registre Y^2 .

Il faudrait donc que la valeur contenue dans Y^2 soit plus grande que 2^{L-1} pour que ϵ^2_0 ait un poids plus grand ou égal à 1.



ce cas est évidemment limité

Dans le cas général $\epsilon^2_0 2^{q_1} h < 2^{q_1} h$

Les termes en ϵ^2_0 sont donc négligeables, puisque de plus $0 < \lambda^1_j - Y^1_j < 1$, on peut écrire :

$$\epsilon^1_i = \epsilon^1_0 (1-ah)^i + 2^{q_1} ah \left[F(R^1_{i-1}) - k_1 \right] + 2^{q_1+q_2} h \left[F(R^2_{i-1}) - k_2 \right] - 2^{q_1} h (1 - (1-ah)^{i-1}) \left[a(F(R^1_j) - k_1) + 2^{q_2} (F(R^2_j) - k_2) \right]$$

III-A-3 Propagation des arrondis au cours d'un pas de calcul

L'arrondi au pas $i-1$ relatif à l'intégrateur ① est

$$a_1(i-1) = h 2^{q_1} \left[F(R^1_{i-1}) - k_1 \right]$$

l'arrondi au pas $i-1$ relatif au multiplieur ② est

$$a_2(i-1) = h 2^{q_1} \left[F(R^2_{i-1}) - k_2 \right]$$

S'il n'y avait pas d'arrondi le registre Y^1 de l'intégrateur recevrait au pas i la valeur

$$h 2^{q_1 + q_2} \left[R^2_{i-1} + Y^2_0 F(R^1_{i-1}) \right]$$

en fait il reçoit la valeur

$$h 2^{q_1 + q_2} E (R^2_{i-1})$$

l'erreur d'arrondi $\epsilon^1_a (i)$ vaut donc

$$\epsilon^1_a (i) = a_2 (i-1) + a \cdot a_1 (i-1)$$

III-A-4 Nouvelle expression de l'erreur de principe. Erreur moyenne.

Comme dans les études précédentes l'erreur de principe peut être notée

$$\begin{aligned} \epsilon^1 (x_i) &= \epsilon^1_a (i) + p^1 (x_i) \\ \text{avec } p^1 (x_i) &= \epsilon^1_o e^{-\alpha(x_i-x_o)} - 2^{q_1} h (1-e^{-\alpha(x_{i-1}-x_o)}) \left[a(F(R^1_j) - k_1) \right. \\ &\quad \left. + 2^{q_2} [F(R^2_j) - k_2] \right] \end{aligned}$$

$$\text{où } \alpha = \frac{L(1-ah)}{h}$$

$$\text{et } - 2^{q_1} h [a k_1 + 2^{q_2} k_2] \leq \epsilon^1_a (i) \leq 2^{q_1} h [a(1-k_1) + 2^{q_2}(1-k_2)]$$

(en supposant k_1 et k_2 positifs)

La courbe de propagation ne diverge pas. On remarquera l'importance des facteurs de cadrage q_1 et q_2 qui déterminent le domaine de variation de l'erreur d'arrondi.

La courbe d'erreur moyenne a pour équation

$$\begin{aligned} \overline{\epsilon^1 [x_o, x_i]} &\neq \frac{1-e^{-\alpha(x_i-x_o)}}{\alpha(x_i-x_o)} \left[\epsilon^1_o + 2^{q_1} h (a [F(R^1_j) - k_1] \right. \\ &\quad \left. + 2^{q_2} [F(R^2_j) - k_2] \right] \end{aligned}$$

III - B Vérification numérique

L'expérimentation avait deux buts :

- 1) Vérifier l'exactitude de l'expression de $\epsilon^1(x_i)$ en faisant varier les valeurs k_1 et k_2

Nous avons fait l'hypothèse

$$\overline{F(R^1_j)} = \overline{F(R^2_j)} = \frac{1}{2}$$

- 2) Tester l'hypothèse $\overline{F(R^1_j)} = \overline{F(R^2_j)} = \frac{1}{2}$

En construisant les courbes d'erreurs moyennes d'équation théorique

$$\overline{\epsilon}_{[x_0, x_i]} = \frac{1-e^{-\alpha(x_i-x_0)}}{\alpha(x_i-x_0)} \left[\epsilon^1_0 + 2^{q_1} h [a(\overline{F(R^1_j)} - k_1) + 2^{q_2} (\overline{F(R^2_j)} - k_2)] \right]$$

Nous avons pris $a = 5, h = 2^{-7}, y_0 = 1, x_0 = 0, q_1 = 0, q_2 = 3, \epsilon_0 = 0$

Remarque : Le fait de prendre une valeur simple pour a , n'a aucune influence sur la démonstration puisque, nous l'avons vu, on peut toujours avoir $\epsilon^2_0 \neq 0$

(A) 1) $k_1 = k_2 = 0$

$$\epsilon^1(x_i) = -\frac{13}{2} h (1-e^{-5,2 x_{i-1}}) + \epsilon^1_a(x_i)$$

avec $0 \leq \epsilon^1_a(x_i) < 13 h$

représentation sur le graphique G.III.1

$$2) \quad k_1 = \frac{1}{2} \quad k_2 = 0$$

$$\epsilon^1(x_i) = -4h(1 - e^{-5,2x_i}) + \epsilon_a^1(x_i)$$

$$-\frac{5}{2} h < \epsilon_a^1(x_i) < \frac{21}{2} h \quad \text{graphique G.III.2}$$

$$3) \quad k_1 = \frac{1}{2}, \quad k_2 = \frac{1}{4}$$

$$\epsilon^1(x_i) = -2h(1 - e^{-5,2x_i}) + \epsilon_a^1(x_i)$$

$$-\frac{9}{2} h < \epsilon_a^1(x_i) < \frac{15}{2} h \quad \text{graphique G.III.3}$$

$$4) \quad k_1 = k_2 = \frac{1}{2}$$

$$\epsilon^1(x_i) = \epsilon_a^1(x_i) \quad \text{avec} \quad -\frac{13}{2} h < \epsilon_a(x_i) < \frac{13}{2} h$$

graphique G.III.4

$$5) \quad k_1 = \frac{1}{2} \quad k_2 = \frac{3}{4}$$

$$\epsilon^1(x_i) = 2h(1 - e^{-5,2x_i}) + \epsilon_a^1(x_i)$$

$$-\frac{15}{2} h < \epsilon_a^1(x_i) < \frac{9}{2} h$$

graphique G III. 5

L'expression donnée de $\epsilon^1(x_i)$ est donc correcte. Lorsque K_1 et K_2 varient, ces variations ont une influence sur l'allure de la courbe de propagation et les limites de variation de l'erreur d'arrondi. L'erreur de principe varie cependant toujours dans des bandes de

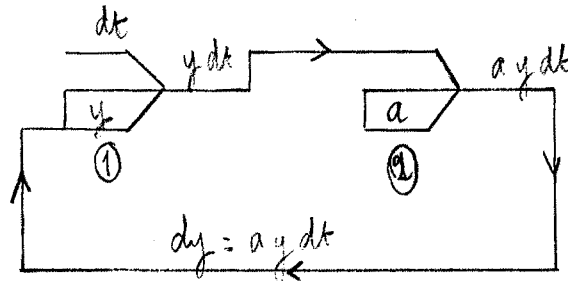
de largeur totale $2^{q_1} h [a + 2^{q_2}] = 13 h$

B) Les courbes d'erreurs moyennes sont représentées sur le graphique G III 6

On peut considérer l'hypothèse $\overline{F(R^1_j)} = \overline{F(R^2_j)} = \frac{1}{2}$ comme acceptable tout au moins lorsque le nombre de pas d'intégrations est assez grand.

L'hypothèse sera d'autant plus rapidement vérifiée que le pas d'intégration h sera petit (c'est le cas des calculs réels où $h \ll 2^{-7}$)

ETUDE DE $y' = + ay$ $a > 0$



IV-A-1 Expression théorique de l'erreur de principe

Le contenu du registre Y^1 au pas i vaut

$$Y^1_i = Y^1_{i-1} + h 2^{q_2} Y^2_o Y^1_{i-1} - h (2^{q_2} Y^2_o \lambda^1_{i-1} + 2^{q_2} \lambda^2_{i-1})$$

En comparant à la relation

$$y^1_i = y^1_{i-1} + ah y^1_{i-1}$$

l'erreur de principe au pas i s'écrit (en ne tenant pas compte des termes en ϵ^2_o)

$$\epsilon^1_i = \epsilon^1_o (1+ah)^{i-2} 2^{q_1} h \left[a (F(R^1_{i-1}) - k_1) + 2^{q_2} (F(R^2_{i-1}) - k_2) \right] + 2^{q_1} h (1 - (1+ah)^{i-1}) \left[a (F(R^1_j) - k_1) + 2^{q_2} (F(R^2_j) - k_2) \right]$$

en passant à la variable $x_i = x_o + ih$ et en posant $\alpha = \frac{L(1+ah)}{h}$

$$\epsilon^1(x_i) = - 2^{q_1} h \left[a (F(R^1_{i-1}) - k_1) + 2^{q_2} (F(R^2_{i-1}) - k_2) \right] \left\{ \epsilon^1_a(x_i) + \epsilon^1_o e^{\alpha(x_i - x_o)} + 2^{q_1} h (1 - e^{\alpha(x_{i-1} - x_o)}) (a (F(R^1_j) - k_1) + 2^{q_2} (F(R^2_j) - k_2)) \right\} P(x_i)$$

et la valeur moyenne dans l'intervalle $[x_0, x_i]$

$$\overline{\epsilon^1}_{[x_0, x_i]} = \frac{1-e^{\alpha(x_i-x_0)}}{\alpha(x_i-x_0)} \left[2^{q_1} h a(F(R^1_j)-k_1) + 2^{q_2} (F(R^2_j)-k_2) - \epsilon^1_0 \right]$$

IV-A-2 Vérification numérique

$$a=+5, h=2^{-9}, q_1=q_2=3; y^1_0=1, x_0=0, \epsilon^1_0 = 0$$

Cas ① $k_1 = k_2 = 0$

$$\epsilon(x_i) = 13.2^{-7} (1-e^{4,9x_i}) + \epsilon_a(x_i)$$

$$13.2^{-6} < \epsilon_a(x_i) < 0 \text{ graphique G.IV}_1$$

Cas ② $k_1 = k_2 = \frac{1}{2}$

$$-13.2^{-7} < \epsilon_a(x_i) < 13.2^{-7}$$

mauvaise vérification lorsque x croît pour la même raison que dans le cas $y' = +y$ (cf II.2) graphique G.IV.2

Valeurs moyennes :

Cas 1 $\overline{\epsilon^1}_{[x_0, x_i]} = \frac{1-e^{4,9 x_i}}{4,9 x_i} \cdot 13.2^{-7}$

Cas 2 $\overline{\epsilon^1}_{[x_0, x_i]} = 0$

graphique G.IV.3

IV-B Multiplication par un nombre négatif (a < 0)

Supposons qu'une valeur a négative soit mise dans le registre Y^2 du multiplieur (Dans l'analyseur différentiel digital étudié, un nombre négatif a pour représentation le complément à 4 de sa valeur absolue)

Le registre R^2 contient des valeurs négatives ($-2 < R^2_j \leq 0$)

et il déborde au pas j si $R^2_j < -1$

Nous devons alors faire l'hypothèse

$$\overline{F(R_j)} [x_0, x_i] = -\frac{1}{2}$$

Pour éliminer le terme de propagation, la valeur K_2 à mettre initialement dans le registre R^2 devra être négative.

Remarque : Les circuits représentatifs de $y' = -|a|y$ et de $y' = +ay$ (avec $a < 0$) ne sont pas strictement équivalents car si R^2 contient des valeurs positives (cas de $y' = -|a|y$), il déborde pour $R^2_j > 1$ et $R^2_j = 1$

Vérification numérique

$a = -5$, circuit $y' = +ay$, $q_1 = 0$, $q_2 = 3$, $h = 2^{-7}$, $y_0 = 1$,
 $x_0 = 0$, $\epsilon_0 = 0$

Nous avons vérifié que :

- Pour éliminer le terme de propagation, il fallait mettre la valeur $k_2 = -\frac{1}{2}$ dans R^2 initialement (par comparaison des courbes d'erreurs moyennes ① et ② sur le graphique G.IV 4)

- Les circuits $y' = -|a|y$ et $y' = +ay$ n'étaient pas équivalents (par comparaison des courbes d'erreurs moyennes ② du graphique G.IV.4 et ④ du graphique G.III.6)

Cependant la précision obtenue est du même ordre de grandeur.

PREMIERES CONCLUSIONS

Ces premières études nous ont montré que :

* L'erreur due à la propagation des erreurs d'arrondi pouvait et devait (car elle peut dériver) être éliminée en mettant des valeurs initiales dans les registres R.

* Lorsque l'erreur de propagation était éliminée, l'étude de l'erreur de principe revenait à une étude de l'erreur d'arrondi.

L'erreur d'arrondi sur le contenu d'un registre Y a pour origine les arrondis sur les incréments que reçoit ce registre.

Ces arrondis se composent suivant certaines lois propres au schéma représentatif du problème.

- Nous avons généralisé la notion d'arrondi sur un incrément.

- Si $h 2^1$ est le poids de l'incrément $E (R^1_j)$

- Si k_1 est la valeur mise initialement dans le registre R^1

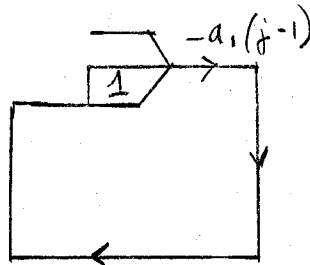
l'arrondi au pas j , $a_1 (j)$ a pour valeur

$$a_1 (j) = h 2^{P_1} (F (R^2_j) - k_1)$$

- Nous avons établi les lois de composition d'arrondi suivantes

α) $y' = + y$

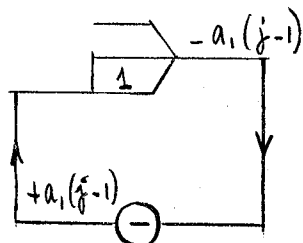
intégration par rapport à la variable indépendante



$$\epsilon^1_a (j) = - a_1 (j - 1)$$

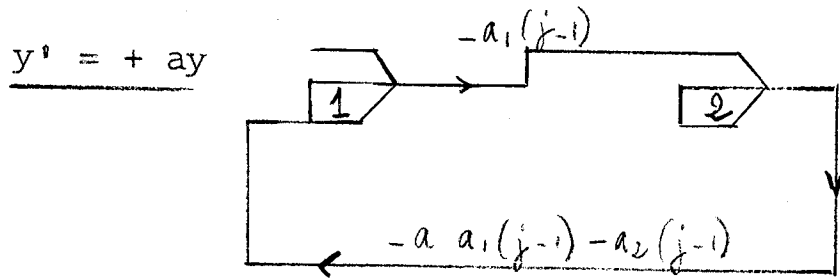
β) $y' = - y$

intégration par rapport à la variable indépendante



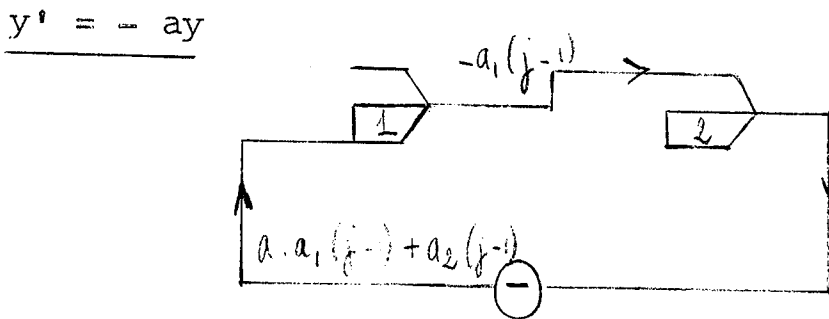
$$\epsilon^1_a (j) = - [-a_1 (j-1)] = a_1(j-1)$$

γ) intégration multiplication par a



$$e^1_a(j) = -a \cdot a_1(j-1) - a_2(j-1)$$

δ) intégration multiplication par a



$$e^1_a(j) = - \left[-a \cdot a_1(j-1) - a_2(j-1) \right]$$

$$= a \cdot a_1(j-1) + a_2(j-1)$$

Nous allons maintenant

- placer deux intégrateurs en chaîne
- réaliser une intégration au sens de Stieljes

Les opérateurs essentiels de l'analyseur digital auront alors été étudiés (le comparateur qui réalise une opération logique et n'utilise qu'un seul registre ne génère pas d'arrondi)

Nous essaierons de déterminer dans quelles conditions ces premières conclusions sont valables pour un circuit quelconque.

C H A P I T R E V

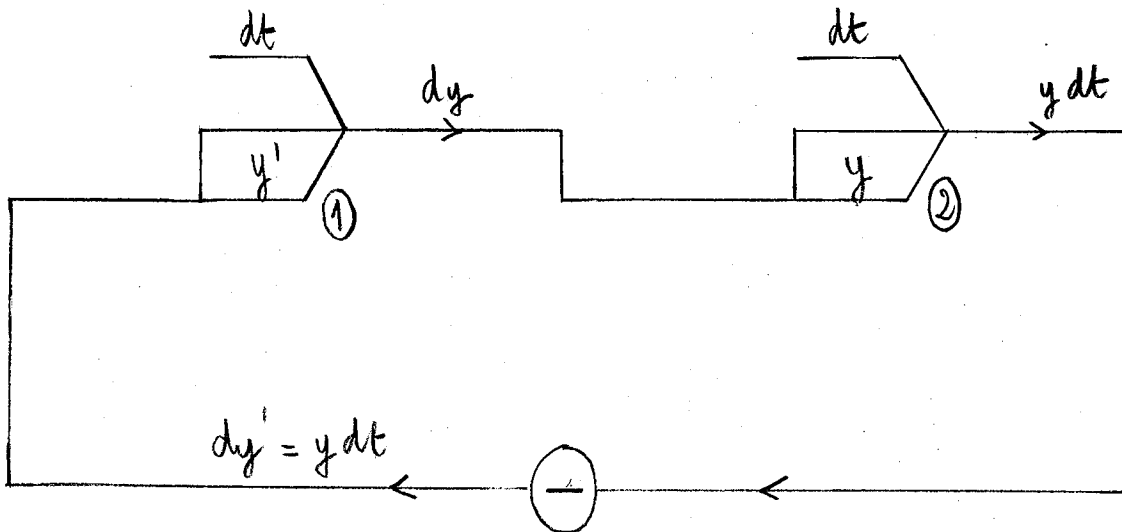
RESOLUTION DE $y'' = -y$.

Si la méthode de résolution est celle de la tangente, le système à représenter est :

$$y_i = y_{i-1} + h y'_{i-1}$$

$$y'_{i-1} = y'_{i-1} + h y''_{i-1} = y'_{i-1} - h y_{i-1}$$

A ce système correspond le schéma suivant :



Nous poserons $y'_i = y^1_i$ et $y_i = y^2_i$

V. A - Phénomène de retard.

Dans l'introduction, nous avons ébauché une étude du phénomène de retard.

Nous avons montré qu'un incrément émis pouvait être utilisé au même cycle d'intégration ou au cycle suivant selon le numéro d'ordre de l'opérateur émetteur.

Dans le circuit que nous étudions actuellement, au pas 0, un incrément $E(R^1_0)$ est émis par l'intégrateur ① ; cet incrément représente $h y^1_0$ soit $\int_{t_0}^{t_1} y^1(u) du$; il est utilisé au même pas pour modifier le registre Y^2

Ce registre contient alors une valeur (Y^2_0) représentative de y^2_1 ; l'incrément émis $E(R^2_0)$ représente donc $h y^2_1$.

L'incrément - $E(R^2_0)$ modifie au pas suivant le registre Y^1 dont le contenu passe de la valeur Y^1_0 à la valeur Y^1_1

A ce niveau, nous avons fait le calcul

$$y^2_1 - y^2_0 = h y^1_0 \quad (\text{correct})$$

$$y^1_1 - y^1_0 = -h y^2_1 \quad (\text{incorrect})$$

et, d'autre part, à la fin du cycle d'intégration 0, les registres Y^1 et Y^2 ont des contenus représentatifs de valeurs à indices différents.

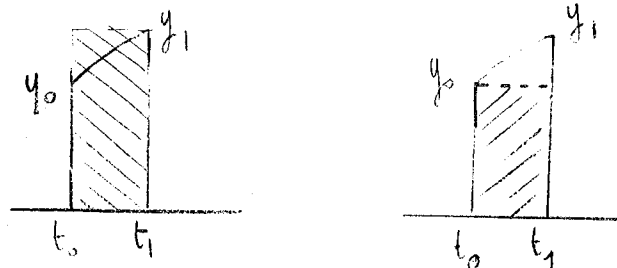
Y^1 contient Y^1_0 représentatif de y^1_0

Y^2 contient (Y^2_0) représentatif de y^2_1

(nous avons déjà rencontré ce phénomène au chapitre I, par. I.C dans l'étude de $y' = -y + f(t)$)

Dans le cas où on intègre par la méthode des rectangles, il est en fait à peu près équivalent de représenter $\int_{t_0}^{t_1} y(u) du$

par $h y_0$ ou par $h y_1$



et le calcul $y^1_1 - y^1_0 = -h y^2_1$ peut-être accepté.

Cette assimilation n'est plus possible lorsqu'on utilise une méthode des trapèzes.

Remarquons tout d'abord que :

$$h (y_{i-1} + \frac{1}{2} (y_{i-1} - y_{i-2}))$$

et
$$h (y_i - \frac{1}{2} (y_i - y_{i-1}))$$

sont des approximations à peu près équivalentes de
$$\int_{t_i}^{t_{i+1}} y(u) du$$

Monsieur VIVIER dans une étude sur les phénomènes de retard distingue :

- Les opérateurs de type 1 qui reçoivent des incréments d'indice $i - 1$ et émettent des incréments d'indice i (cas de l'intégrateur ①)

- Les opérateurs de type 0 qui reçoivent des incréments d'indice i et émettent des incréments d'indice i (cas de l'intégrateur ②)

Au cycle de calcul i , les valeurs représentées par les contenus des registres Y passent :

- . De l'indice $i - 1$ à l'indice i pour les intégrateurs de type 1
- . De l'indice i à l'indice $i + 1$ pour les intégrateurs de type 0.

Dans notre circuit :

$$Y_i^1 \leftrightarrow 2^{-q_1} Y_i^1 \quad Y_i^2 \leftrightarrow 2^{-q_2} Y_{i+1}^2 \quad (1)$$

et les formules à représenter sont donc :

$$Y_i^2 - Y_{i-1}^2 = h (Y_{i-1}^1 + \frac{1}{2} (Y_{i-1}^1 - Y_{i-2}^1)) \text{ type 1}$$

$$Y_i^1 - Y_{i-1}^1 = -h (Y_i^2 - \frac{1}{2} (Y_i^2 - Y_{i-1}^2)) \text{ type 0}$$

Cependant, le fait que les registres Y contiennent des valeurs à indices différentes est un inconvénient en cas d'extraction en fin de cycle.

Monsieur VIVIER préconise de ne faire fonctionner que certains intégrateurs au cycle 0.

Supposons en effet que seul l'intégrateur (2) fonctionne au cycle 0. Cet intégrateur émettra $E(R^2_0)$ représentatif de $h y^1_0$

Au cycle 1, on fera le calcul

$$y^1_1 - y^1_0 = - h y^1_0$$

$$\boxed{y^2_1 - y^2_0 = + h y^1_1}$$

et à la fin du cycle i , les contenus des registres Y représenteront des valeurs de même indice.

$$Y^1_i \iff 2^{-q_1} y^1_i \quad Y^2_i \iff 2^{-q_2} y^2_i$$

mais les intégrateurs changeront de type

$$y^2_i - y^2_{i-1} = h (y^1_i - \frac{1}{2} (y^1_i - y^1_{i-1})) \quad \text{type 0}$$

$$y^1_i - y^1_{i-1} = - h (y^2_{i-1} + \frac{1}{2} (y^2_{i-1} - y^2_{i-2})) \quad \text{type 1}$$

Il peut y avoir une certaine difficulté technologique à faire démarrer seulement certains opérateurs au cycle 0. Nous verrons qu'il existe une solution très simple.

Dans la première partie de ce chapitre, nous n'allons pas éliminer le phénomène de retard. Nous utiliserons une méthode d'intégration des rectangles.

Le calcul débutera à l'intégrateur (1) ;

et au pas i nous comparerons les valeurs Y^1_i et Y^2_i aux solutions du système

$$\begin{aligned} y^1_i &= y^1_{i-1} - h y^2_{i-1} \\ y^2_i &= y^2_{i-1} + h y^1_{i-1} \end{aligned} \quad (1)$$

Nous faisons donc délibérément apparaître une erreur de retard ; en effet :

* au pas i , Y_i^2 est représentatif de la valeur y_{i+1}^2 et non de y_i^2 .

* Le système effectivement représenté sur l'ADD n'est pas le système (1) mais :

$$y_i^1 = y_{i-1}^1 - h y_{i-1}^2$$

$$y_i^2 = y_{i-1}^2 + h y_{i-1}^1$$

Nous étudierons d'autre part :

- La nouvelle loi de composition des arrondis
- La courbe de propagation des erreurs d'arrondis

Ensuite, dans une deuxième partie, nous utiliserons une méthode des trapèzes.

Nous éliminerons le retard en faisant démarrer le calcul à l'intégrateur (2) et en intégrant avec 2 formules.

En éliminant le terme de propagation des arrondis, nous vérifierons (nous avons déjà fait cette vérification en I.D) que l'erreur de principe vaut bien l'erreur d'arrondi.

V- B ETUDE DE L'ERREUR DE PRINCIPE DANS LE CAS D'UNE INTEGRATION PAR LA METHODE DES RECTANGLES.

V-B.1 Relations fondamentales

$$\begin{aligned} Y^1_i &= Y^1_{i-1} - h 2^{q_2 - q_1} & Y^2_{i-1} &+ h 2^{q_2 - q_1} \lambda^2_{i-1} \\ Y^2_i &= Y^2_{i-1} + h 2^{q_1 - q_2} & Y^1_i &- h 2^{q_1 - q_2} \lambda^1_i \end{aligned}$$

au départ, les registres Y^1 et Y^2 contiennent respectivement Y^1_0 et Y^2_0

V-B.2 Expression théorique de l'erreur de principe

De la même manière que dans les études précédentes, on peut déterminer au pas i , les erreurs de principe ϵ^1_i et ϵ^2_i

$$\epsilon^1_i = 2^{q_1} Y^1_i - y^1_i \quad \text{et} \quad \epsilon^2_i = 2^{q_2} Y^2_i - y^2_i$$

En comparant les relations fondamentales et le système

$$y^1_i = y^1_{i-1} - h y^2_{i-1}$$

$$y^2_i = y^2_{i-1} + h y^1_{i-1}$$

on obtient le système aux différences finies suivant

$$\epsilon^1_i = \epsilon^1_{i-1} - h \epsilon^2_{i-1} + h 2^{q_2} \lambda^2_{i-1}$$

$$\epsilon^2_i = \epsilon^2_{i-1} + h \epsilon^1_{i-1} - h 2^{q_1} \lambda^1_i + h 2^{q_1} (Y^1_i - Y^1_{i-1})$$

système qui peut s'écrire sous la forme

$$Y_i = A Y_{i-1} + C_{i-1} \quad (1)$$

où $Y_i = \begin{pmatrix} \epsilon^1_i \\ \epsilon^2_i \end{pmatrix}$ $A = \begin{pmatrix} 1 & -h \\ h & 1 \end{pmatrix}$

Soient α_1 et α_2 les valeurs propres de A données par l'équation caractéristique $(1 - \alpha)^2 + h^2 = 0$

Soit $X_1 = \begin{pmatrix} 1 \\ -j \end{pmatrix}$ un vecteur propre correspondant à

$$\alpha_1 = 1 + jh$$

$X_2 = \begin{pmatrix} 1 \\ +j \end{pmatrix}$ un vecteur propre correspondant à

$$\alpha_2 = 1 - jh$$

Multiplions à gauche la relation matricielle (1) par X_1^T

$$X_1^T \cdot Y_i = X_1^T \cdot A \cdot Y_{i-1} + X_1^T \cdot C_{i-1}$$

Par définition de α_1 , $A X_1 = \alpha_1 X_1$

$$\text{et } X_1^T \cdot A^T = \alpha_1 \cdot X_1^T$$

Imaginons alors de rechercher les valeurs propres et les vecteurs propres de A^T .

Les valeurs propres seront toujours α_1 et α_2 , mais à α_1 correspondra le vecteur propre X_2 et à α_2 correspondra le vecteur propre X_1

$$\text{Ainsi } A^T \cdot X_1 = \alpha_2 X_1$$

et donc

$$X_1^T (A^T)^T = \alpha_2 \cdot X_1^T$$

$$X_1^T \cdot A = \alpha_2 \cdot X_1^T$$

Ainsi

$$X_1^T \cdot Y_i = \alpha_2 X_1^T Y_{i-1} + X_1^T C_{i-1}$$

De même

$$X_2^T \cdot Y_i = \alpha_1 X_2^T Y_{i-1} + X_2^T C_{i-1}$$

On peut ainsi écrire

$$X^T_1 Y_i = \alpha^i_2 X^T_1 Y_0 + \sum_{j=0}^{i-1} \alpha^{i-1-j}_2 X^T_1 C_j$$

$$X^T_2 Y_i = \alpha^i_1 X^T_2 Y_0 + \sum_{j=0}^{i-1} \alpha^{i-1-j}_1 X^T_2 C_j$$

D'une manière plus explicite

$$\epsilon^1_i - j \epsilon^2_i = \alpha^i_2 (\epsilon^1_0 - j \epsilon^2_0)$$

$$+ h 2^{q_2} \sum_0^{i-1} \alpha^k_2 \lambda^2_{i-k-1} + j h 2^{q_1} \sum_0^{i-1} \alpha^k_2 \lambda^1_{i-k}$$

$$- j h 2^{q_1} \sum_0^{i-1} \alpha^k_2 (Y^1_{i-k} - Y^1_{i-k-1})$$

De même :

$$\epsilon^2_i + j \epsilon^1_i = \alpha^i_1 (\epsilon^1_0 + j \epsilon^2_0)$$

$$+ h 2^{q_2} \sum_0^{i-1} \alpha^k_1 \lambda^2_{i-k-1} - j h 2^{q_1} \sum_0^{i-1} \alpha^k_1 \lambda^1_{i-k}$$

$$+ j h 2^{q_1} \sum_0^{i-1} \alpha^k_1 (Y^1_{i-k} - Y^1_{i-k-1})$$

On a ainsi :

$$2 \epsilon^1_i = \epsilon^1_0 (\alpha^i_2 + \alpha^i_1) - j \epsilon^2_0 (\alpha^i_2 - \alpha^i_1)$$

$$+ h 2^{q_2} \sum_0^{i-1} (\alpha^k_2 + \alpha^k_1) \lambda^2_{i-k-1}$$

$$+ j h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 - \alpha^k_1) \lambda^1_{i-k}$$

$$- j h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 - \alpha^k_1) (Y^1_{i-k} - Y^1_{i-k-1})$$

et

$$\begin{aligned}
 -2 j \epsilon^2_i &= \epsilon^1_o (\alpha^i_2 - \alpha^i_1) - j \epsilon^2_o (\alpha^i_2 + \alpha^i_1) \\
 &+ h 2^{q_2} \sum_0^{i-1} (\alpha^k_2 - \alpha^k_1) \lambda^2_{i-k-1} + j h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 + \alpha^k_1) \lambda^1_{i-k} \\
 &- j h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 + \alpha^k_1) (Y^1_{i-k} - Y^1_{i-k-1})
 \end{aligned}$$

Ces expressions ne sont pas complètes.

Au pas 0, en effet le registre Y^2 de l'intégrateur (2) est modifié par l'incrément émis par l'intégrateur (1)

Le registre Y^2 contient, non pas la valeur initiale Y^2_o mais une valeur

$$(Y^2_o) = Y^2_o + h 2^{q_1} Y^1_o - h 2^{q_1} - 2^{q_2} \lambda^1_o$$

On peut considérer que l'erreur initiale est

$$(\epsilon^2_o) = \epsilon^2_o + h 2^{q_1} Y^1_o - h 2^{q_1} \lambda^1_o$$

$$\text{avec } \lambda^1_o = F(R^1_o) - k_1$$

(k_1 , valeur mise initialement dans le registre R^1)

Les deux relations précédentes doivent donc s'écrire

$$\begin{aligned}
 2 \epsilon^1_i &= \epsilon^1_o (\alpha^i_2 + \alpha^i_1) - j \epsilon^2_o (\alpha^i_2 - \alpha^i_1) \\
 &+ h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 + \alpha^k_1) \lambda^2_{i-k-1} + j h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 - \alpha^k_1) \lambda^1_{i-k} \\
 &- j h 2^{q_1} \sum_0^{i-1} (\alpha^k_2 - \alpha^k_1) (Y^1_{i-k} - Y^1_{i-k-1})
 \end{aligned}$$

$$\begin{aligned} - 2j \epsilon_i^2 &= \epsilon_o^1 (\alpha_2^i - \alpha_1^i) - j \epsilon_o^2 (\alpha_2^i + \alpha_1^i) \\ &+ h_2^{q_2} \sum_0^{i-1} (\alpha_2^k - \alpha_1^k) \lambda_{i-k-1}^2 + j h_2^{q_1} \sum_0^i (\alpha_2^k + \alpha_1^k) \lambda_{i-k}^1 \\ &- j h_2^{q_1} \sum_0^i (\alpha_2^k + \alpha_1^k) (Y_{i-k}^1 - Y_{i-k-1}^1) \end{aligned}$$

V. B. 3 Etude de ϵ_i^2

Utilisons les variables $F(R_K^1)$ et $F(R_K^2)$ et notons

$$\overline{F(R_K^1)}_{(0,j)} = \beta_1(j)$$

$$\overline{F(R_K^2)}_{(0,j)} = \beta_2(j)$$

Comme le pas $h \ll 1$, on peut assimiler $\alpha_1^K + \alpha_2^K = (1+jh)^K + (1-jh)^K$

à $2\cos Kh$; de même $\alpha_1^K - \alpha_2^K \approx -2j \sin Kh$.

(Ces relations sont vraies à un terme multiplicatif près $|1 + jh|^K$; nous tiendrons compte de ce terme en V.B.4)

ϵ_i^2 peut alors s'écrire :

$$\begin{aligned} \epsilon_i^2 &= \epsilon_0^1 \sin ih + \epsilon_0^2 \cos ih && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_{po}^2(i) \\ &+ h 2^{q_2} (\beta_2(i-1) - k_2) \sin (i-1)h && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_{pa}^2(i) \\ &+ h 2^{q_1} (\beta_1(i-1) - k_1) (1 - \cos ih) && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_{pr}^2(i) \\ &+ h 2^{q_1} \sum_1^{i-1} \cos Kh (Y_{i-K}^1 - Y_{i-K-1}^1) && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_a^2(i) \\ &+ h 2^{q_1} \cos ih \cdot Y_0^1 && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_r^2(i) \\ &- h 2^{q_1} (F(R_i^1) - k_1) && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_a^2(i) \\ &+ h 2^{q_1} (Y_i^1 - Y_{i-1}^1) && \left. \vphantom{\epsilon_i^2} \right\} \epsilon_r^2(i) \end{aligned}$$

Remarques :

a) Loi de composition des arrondis

$$\epsilon_a^2(i) = -h 2^{q_1} [F(R_i^1) - k_1] = -a_1(i)$$

Nous avons défini, dans l'introduction l'erreur d'arrondi comme une fonction des arrondis au pas i et au pas $i - 1$.

La valeur de $\epsilon_a^2(i)$ devrait donc être

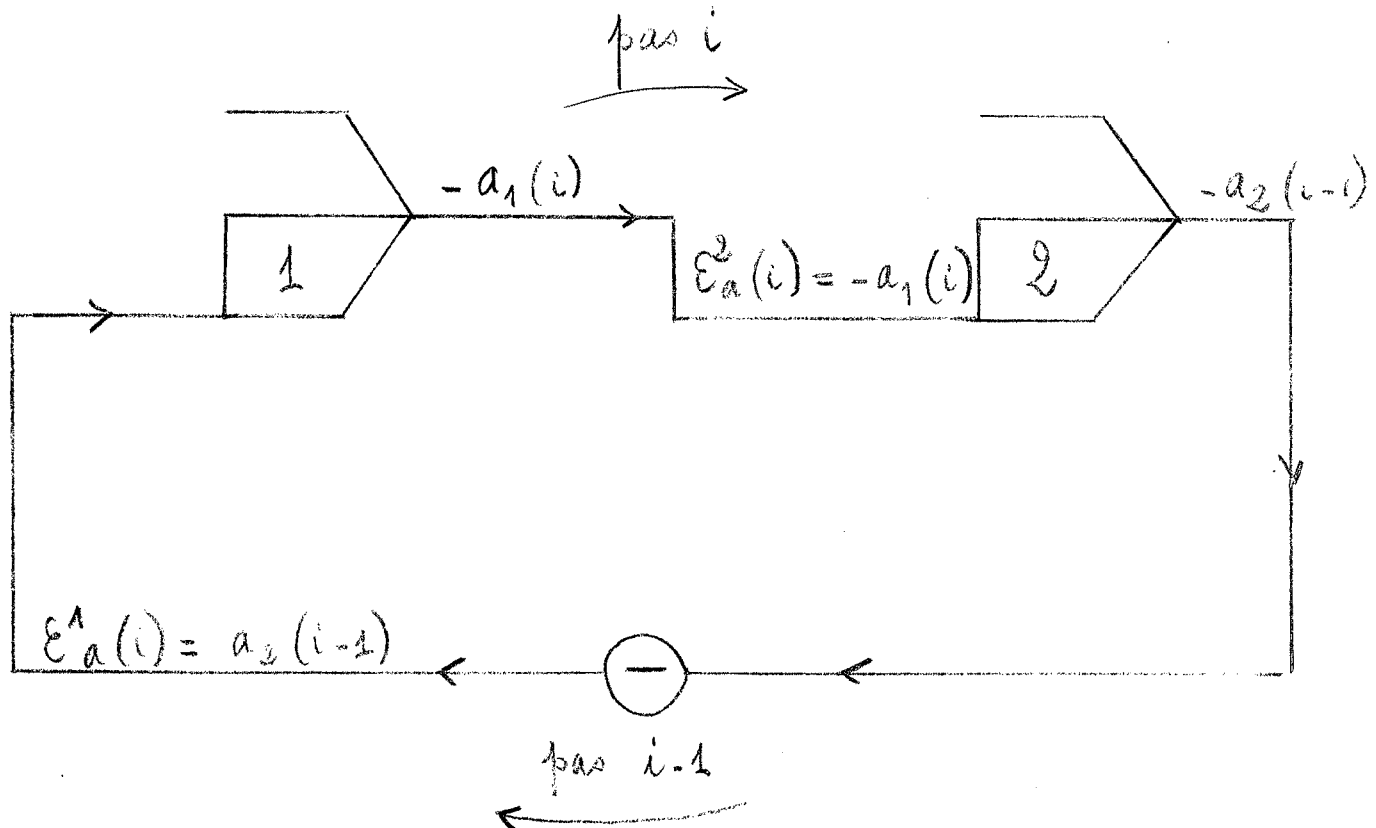
$$\epsilon_a^2(i) = -a_1(i) + h a_2(i-1)$$

La valeur $h a_2(i-1)$ étant faible devant $a_1(i)$, nous avons préféré inclure ce terme dans $\epsilon_{pa}(i)$

On démontrait de même que

$$\epsilon_a^1(i) = h 2^{q_1} [F(R_{i-1}^2) - k_2] = +a_2(i-1)$$

d'où le schéma



β) Etude du terme $\epsilon^2_{pr}(i)$ (propagation des erreurs de retard)

$$\epsilon^2_{pr}(i) = h 2^{q_1} \sum_1^{i-1} \cos Kh (Y^1_{i-K} - Y^1_{i-K-1}) + h 2^{q_1} Y^1_0 \cos ih$$

Remarquons que le terme $h 2^{q_1} Y^1_0 \cos x_i$ peut être facilement éliminé de la manière suivante :

au lieu de mettre une valeur k_1 initialement dans le registre R^1 , mettons une valeur $k'_1 = k_1 - Y^1_0$.

Dans l'expression de ϵ^2_i , les termes transformés sont $+ h 2^{q_1} (\beta_1(i-1) - k'_1)(1 - \cos ih)$ et $- h 2^{q_1} (F(R^1_i) - k'_1)$

En remplaçant k'_1 par sa valeur

$$\begin{aligned} \epsilon^2_i &= \epsilon^1_0 \sin ih + \epsilon^2_0 \cos ih && \left. \vphantom{\epsilon^2_i} \right\} \epsilon^2_{po}(i) \\ &+ h 2^{q_2} (\beta_2(i-1) - k_2) \sin (i-1)h && \left. \vphantom{\epsilon^2_i} \right\} \epsilon^2_{pa}(i) \\ &+ h 2^{q_1} \left[(\beta_1(i-1) - k_1)(1 - \cos ih) + Y^1_0 (\cancel{1 - \cos ih}) \right] && \left. \vphantom{\epsilon^2_i} \right\} \epsilon^2_{pr}(i) \\ &+ h 2^{q_1} \left[\sum_1^{i-1} \cos Kh (Y^1_{i-K} - Y^1_{i-K-1}) + Y^1_0 \cancel{\cos ih} \right] && \left. \vphantom{\epsilon^2_i} \right\} \epsilon^2_{pr}(i) \\ &- h 2^{q_1} (F(R^1_i) - k_1) + \cancel{Y^1_0} && \left. \vphantom{\epsilon^2_i} \right\} \epsilon^2_a(i) \\ &+ \epsilon^2_r(i) \end{aligned}$$

Nous retrouvons l'expression initiale de ϵ^2_a en éliminant le terme $h 2^{q_1} Y^1_0 \cos ih$ dans $\epsilon^2_{pr}(i)$

Nous verrons que cette remarque a une grande importance lorsque les intégrations sont faites par la méthode des trapèzes.

Dans la méthode des rectangles, le terme $h 2^{q_1} Y_0^1 \cos h$ est de signe contraire à $h 2^{q_1} \sum_{k=1}^{i-1} \cos k h (Y_{i-k}^1 - Y_{i-k-1}^1)$, il ne doit donc pas être éliminé.

γ) Equation de la courbe de propagation

$$\begin{aligned}
 P_2(x_i) &= \epsilon_{po}^2(x_i) + \epsilon_{pa}^2(x_i) + \epsilon_{pr}^2(x_i) \\
 &\left. \begin{aligned}
 &\neq \epsilon_0^1 \sin(x_i - x_0) + \epsilon_0^2 \cos(x_i - x_0) \\
 &+ h 2^{q_2} (\beta_2(x_{i-1} - x_0) - k_2) \sin(x_{i-1} - x_0) \\
 &+ h 2^{q_1} (\beta_1(x_{i-1} - x_0) - k_1) (1 - \cos(x_i - x_0))
 \end{aligned} \right\} \epsilon_{po}^2(x_i) \\
 &\left. \begin{aligned}
 &+ h \frac{\sin(x_i - x_0) \cos(x_i - x_0)}{x_i - x_0}
 \end{aligned} \right\} \epsilon_{pr}^2(x_i)
 \end{aligned}$$

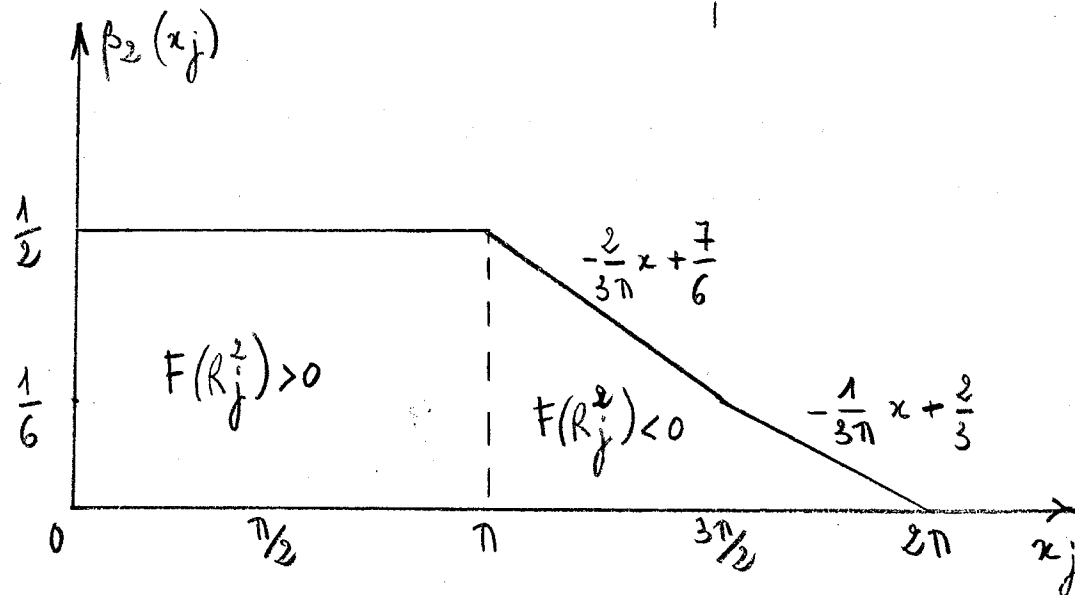
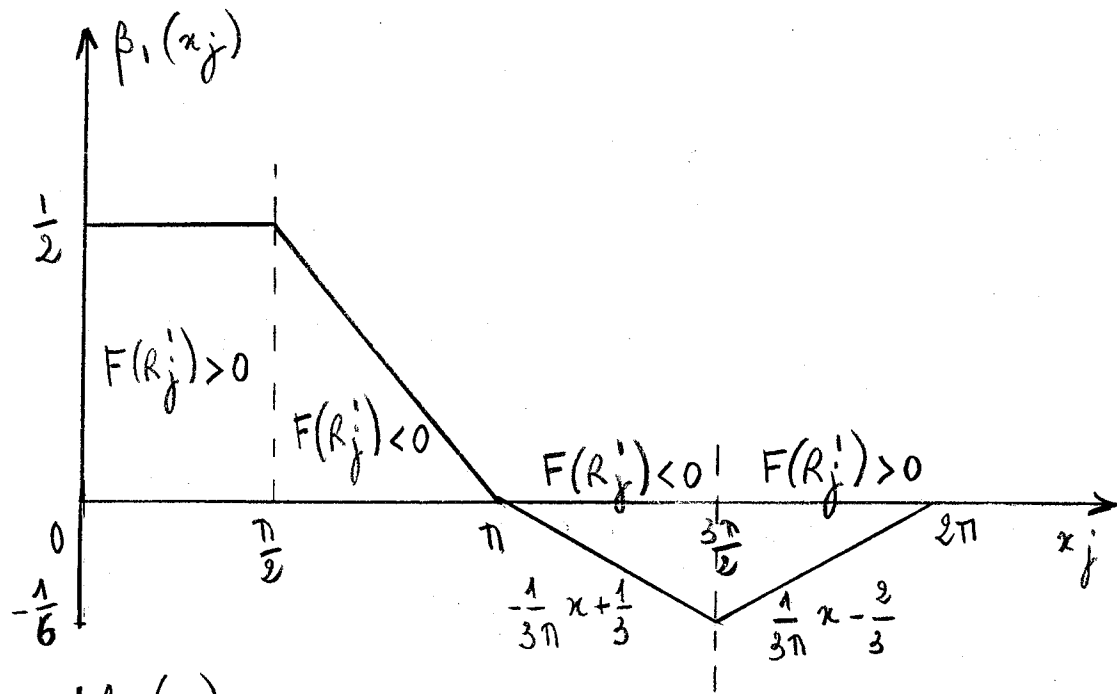
remarques : Supposons que x_i varie de 0 à 2π

a) $\epsilon_{pr}^2(x_i) = \frac{h}{2x_i} \sin 2x_i$

ce terme a donc peu d'influence sur ϵ_i^2 lorsque $x_i > \frac{\pi}{2}$

b) $\beta_1(x_j)$ et $\beta_2(x_j)$ ont des valeurs variables.

Supposons que $y_0^1 = 1$ et $y_0^2 = 0$; on peut estimer que les valeurs moyennes des contenus des registres R^1 et R^2 varient de la manière suivante :



En mettant des valeurs k_1 et k_2 initialement dans les registres R^1 et R^2 , $\epsilon^2_{pa}(x_i)$ ne pourra donc pas être éliminée dans tout l'intervalle d'intégration.

V.B4. Vérification numérique

Cas (1) sur le graphique G.V.1.

$$h = 2^{-5}, y^1_0 = 1, y^2_0 = 0, x_0 = 0 ; \epsilon^1_0 = \epsilon^2_0 = 0 ; q_1 = q_2 = 1 ;$$

$$\boxed{K_1 = K_2 = 0}$$

* La courbe de propagation théorique diffère un peu de la courbe de propagation réelle lorsque x_i croît ; elle a tendance à augmenter d'amplitude.

L'explication de ce phénomène est dans l'assimilation faite en V.B.3 des termes $\alpha^K_1 + \alpha^K_2$ et $\alpha^K_1 - \alpha^K_2$ à $2 \cos k h$ et $- 2 j \sin k h$ respectivement.

Ces relations sont valables à un terme multiplicatif près $|1 + jh|^K$, terme qui augmente avec K.

* Lorsque $x_i = \frac{\pi}{2}$; le terme Y^1_i prend la valeur zéro puis devient négatif. Lorsque des valeurs Y^1_i négatives lui sont ajoutées le contenu du registre R^1 , qui était positif, est remis à zéro puis devient négatif.

$\epsilon^2_a(i) = h 2^{q_1} F(R^1_i)$ change de signe si bien que la relation

$$\epsilon^2_i = p(x_i) + \epsilon^2_a(i) \text{ est toujours vérifiée.}$$

Ce phénomène, qui se traduit par une courbe parabolique sur le graphique est, en grande partie, à l'origine des fluctuations de l'erreur de principe.

Il serait donc intéressant de stabiliser l'erreur de principe à ce niveau. On peut imaginer de faire un calcul par tranches c'est-à-dire d'introduire dans les registres R^1 et R^2 , lorsque $Y^1_i = 0$, des valeurs qui éliminent le terme de propagation des erreurs d'arrondi dans l'intervalle de variation $(\frac{\pi}{2}, \pi)$. Ces valeurs accélèreraient l'émission du premier incrément négatif $E(R^1_i)$ et la variation du type parabolique de ϵ^2_i disparaîtrait.

(Dans l'intervalle $[\frac{\pi}{2}, \pi]$, $F(R^1_j) = -\frac{1}{2}$; il faudrait donc que le registre R^1 contienne $-\frac{1}{2}$ au début de la nouvelle tranche de calcul)

On pourrait donc considérer que le calcul débute en $\frac{\pi}{2}$ avec des erreurs initiales

$$\epsilon^1\left(\frac{\pi}{2}\right) \text{ et } \epsilon^2\left(\frac{\pi}{2}\right).$$

Nous reverrons ce problème de stabilisation d'une manière plus approfondie lorsque nous intégrerons par la méthode des trapèzes.

Cas (2) sur G.V.2 $K_1 = K_2 = 0$ mais $\boxed{\epsilon^2_0 = -h}$

dans ce cas $\epsilon_{po}(x_i) \neq -h \cos x_i$

Cas (3) sur G.V.3 $\boxed{K_1 = K_2 = \frac{1}{2}}$ $\epsilon'_0 = \epsilon^2_0 = 0$

Le terme de propagation des erreurs d'arrondi est nul de 0 à $\frac{\pi}{2}$. Cependant ϵ^2_i ne varie pas de $-h$ à $+h$ car $p(x_i) = \epsilon_{pr}(x_i) \neq 0$

Cas (4) sur G.V.4 $K_1 = -0,1875$ $K_2 = 0,375$

La valeur de $p(x_i)$ est minimisée dans l'intervalle $(\pi, \frac{3\pi}{2})$; $\epsilon_a^2(i) = -h^2 \left[F(R_i^1) - k_1 \right]$ varie de $-0,4 h$ à $1,6 h$ puisque $F(R_i^1)$ est négatif dans cet intervalle.

Lorsque $K_1 = K_2 = \frac{1}{2}$, $\epsilon_a^2(i)$ variait de $+h$ à $+3h$ pour $x_i \in [\pi, \frac{3\pi}{2}]$ si bien que très souvent,

$$\begin{array}{ccc} \left| \epsilon_a^2(i) \right| & < & \left| \epsilon_a^2(i) \right| \\ \text{cas (3)} & & \text{cas (4)} \end{array}$$

Donc, lorsque dans un intervalle d'intégration les registres R contiennent des valeurs à signe variable, la minimisation de $p(x_i)$ n'est pas un critère de minimisation de la valeur absolue de l'erreur de principe.

V.C Integrations par la méthode des trapèzes

En étudiant l'intégration par la méthode des rectangles, nous avons montré que le terme $+ h 2^{q_1} Y^1_0 \cos ih$ intervenant dans l'expression de $\epsilon^2_{pr}(x_i)$, pouvait être éliminé en mettant la valeur

$$k'_1 = k_1 - Y^1_0 \quad \text{dans le registre } R^1$$

Ce terme apparaissait dans l'expression de $\epsilon^2_{pr}(x_i)$ parce que l'on comparait Y^2_i à y^2_i (alors que en fait Y^2_i est représentatif de y^2_{i+1}).

On a en effet :

$$y^2_{i+1} - y^2_i \neq h 2^{q_1} Y^1_0 \cos ih \quad \text{puisque } 2^{q_1} Y^1_0 = 1$$

En mettant k'_1 dans le registre R^1 , tout se passe comme si, au cycle 0, l'intégrateur (1) ne fonctionnait pas. L'erreur de retard est donc éliminée.

Nous allons représenter le système

$$y^1_i = y^1_{i-1} - h (y^2_{i-1} + \frac{1}{2} (y^2_{i-1} - y^2_{i-2}))$$

$$y^2_i = y^2_{i-1} + h (y^1_i - \frac{1}{2} (y^1_i - y^1_{i-1}))$$

et mettre des valeurs $k'_1 = k_1 - Y^1_0$ et k_2 dans les registres R^1 et R^2 .

Nous allons faire les hypothèse suivantes :

lorsque les intégrations sont faites par la méthode des rectangles et que le phénomène de retard est éliminé

$$\epsilon^2_i = \epsilon^2_{po}(i) + \epsilon^2_{pa}(i) + \epsilon^2_a(i)$$

où

- $\epsilon_{pa}^2(i)$ est le terme de propagation des erreurs d'arrondi.

Nous supposons que ce terme a la forme

$$h 2^{q_2} (\beta_2(j) - k_2) \theta_2(j) + h 2^{q_1} (\beta_1(j) - k_1) \theta_1(j)$$

et qu'il est donc éliminable.

- $\epsilon_a^2(i)$ est l'erreur d'arrondi ; nous supposons qu'elle obéit aux lois de composition.

Nous avons vu au paragraphe I.D que ces hypothèses étaient vérifiées dans le cas du problème $y'' = -y$

Soient $h = 2^{-5}$, $y_0^1 = 1$, $y_0^2 = 0$, $\epsilon_0^1 = \epsilon_0^2 = 0$

$q_1 = q_2 = 1$, $k_1 = \frac{1}{2}$; $k'_1 = k_1 - y_0^1 = 0$; $k_2 = \frac{1}{2}$

Si les hypothèses précédentes sont vérifiées, on doit avoir dans l'intervalle $(0, \frac{\pi}{2})$

$$\epsilon_{po}^2(i) = 0$$

$$\epsilon_{pa}^2(i) = 0$$

et donc $\epsilon_i^2 = \epsilon_a^2(i) = -a_1(i) = -h 2^{q_1} \left[F(R_i^1) - \frac{1}{2} \right]$

On peut constater sur le graphique G.V.5 que ϵ_i^2 varie bien dans les limites $(+h, -h)$ lorsque x_i varie de 0 à $\frac{\pi}{2}$.

Sur le graphique G.V.6 est représentée la courbe de variation de ϵ_i^1 ; on voit de même que :

$$\epsilon_i^1 = \epsilon_a^1(i) = -a_2(i-1) = -h 2^{q_2} \left[F(R_{i-1}^2) - \frac{1}{2} \right]$$

Remarque :

Supposons que la valeur mise initialement dans le registre Y_1 soit non pas $k'_1 = 0$ mais $k'_1 = \frac{1}{2}$.

On peut vérifier sur le graphique G.V.7 que la courbe de propagation a bien une équation de la forme $+ h \cos x_i$ (tout au moins dans l'intervalle $(0, \frac{\pi}{2}]$)

V. D Essai de stabilisation

Nous avons vu au paragraphe V.B.4 que le passage de Y^1_i par la valeur zéro était une nouvelle cause d'erreur.

Pour analyser cette cause d'erreur, nous rappellerons la définition de l'arrondi :

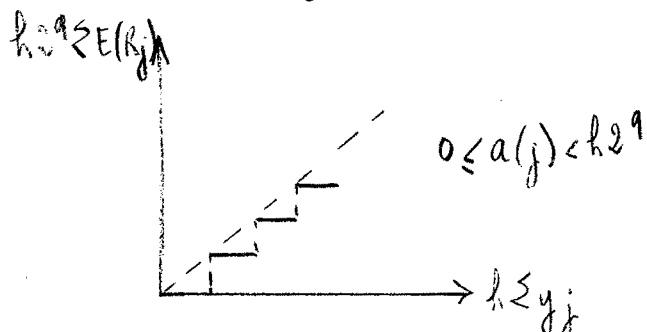
lorsqu'un intégrateur représente la relation

$$z_i = z_0 + h \sum y_j$$

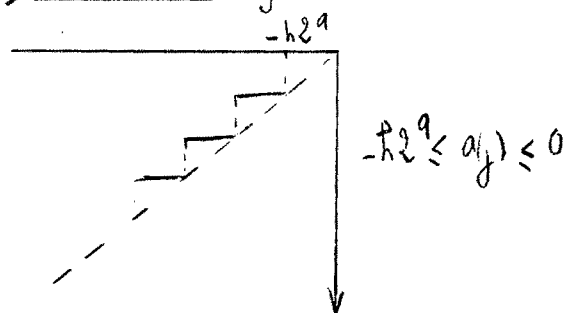
et que les y_j sont connus sans erreur, la valeur de z_i est obtenue avec un arrondi $a(i) = h 2^q [F(R_i) - k]$

Si $k = 0$, nous avons représenté l'arrondi sous deux formes

1) tous les y_j sont > 0



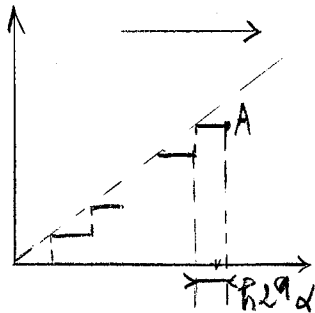
2) tous les y_j sont < 0



Nous n'avons pas étudié le cas où, dans un même calcul, les y_j ont des signes différents.

Dans l'étude de $y'' = -y$, pour $x_i \neq \frac{\pi}{2}$, les intégrandes Y^1_j deviennent négatifs.

1ère phase $\frac{\pi}{2} > \alpha_i > 0$



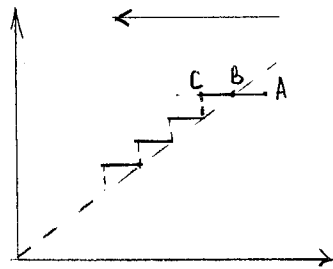
$$0 \leq a(j) < h 2^q \alpha$$

Lorsque $Y^1_j = 0$, $F(R^1_j) \neq 0$ le plus généralement

$$F(R^1_j) = \alpha$$

$$\text{et } a(j) = h 2^q \alpha.$$

2ème phase $\pi > \alpha_i > \frac{\pi}{2}$



$$-h 2^q \leq a(j) \leq 0$$

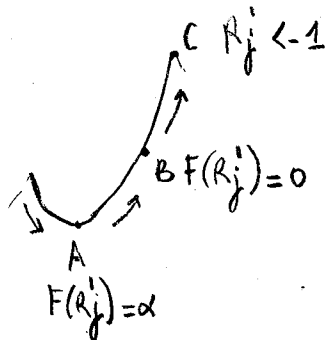
Lorsque Y^1_j devient négatif, le contenu de R passe par la valeur zéro (point B) puis devient négatif.

Un incrément -1 est émis (point C) lorsque $R_i < -1$

sauf au départ (zone AC)

ou $-h 2^q < a(j) < h 2^q \alpha$

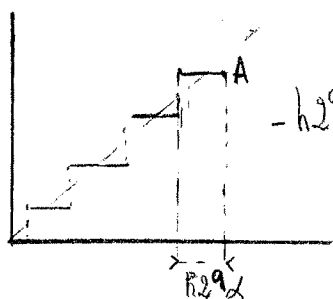
On voit ainsi que l'intervalle de variation de $a(j)$ est anormal au moment où Y^1_j change de signe. On a l'explication de la forme parabolique de la courbe d'erreur de principe.



② $|k| = \frac{1}{2}$

Pour que les arrondis se compensent statistiquement, nous avons cherché à faire varier ces arrondis entre les bornes $-h 2^{q-1}$ et $+h 2^{q-1}$ en mettant une valeur initiale $\frac{1}{2}$ dans le registre R.

1ère phase



$$-h2^{q-1} \leq a(j) < h2^{q-1}$$

en A, $F(R_j^1) = \alpha$

2ème phase

lorsque Y_j^1 devient négatif

$$F(R_j^1) = 0 \text{ en C}$$

et un incrément -1 est émis (en D) lorsque $R_j < -1$

On a donc $-3h2^{q-1} < a(j) < -h2^{q-1}$
sauf au départ où $-3h2^{q-1} < a(j) < h2^q(\alpha - \frac{1}{2})$

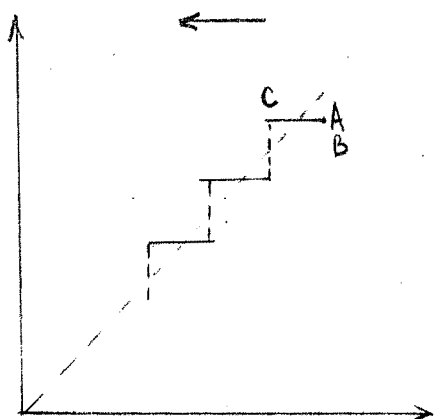
La loi de variation de l'arrondi est très mauvaise.

1ère solution :

Décider d'annuler le contenu du registre R lorsque Y_j passe par la valeur zéro.

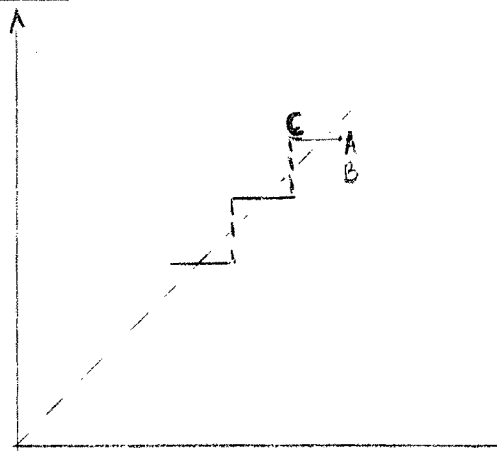
La loi d'arrondi est meilleure

$k=0$; 2^e phase



$$-h2^q(1-\alpha) \leq a(j) \leq h2^q\alpha$$

$k=+\frac{1}{2}$; 2^e phase



$$-h2^q(\frac{3}{2}-\alpha) \leq a(j) \leq h2^q(\alpha - \frac{1}{2})$$

Expérience numérique :

Nous avons utilisé cette méthode et représenté les courbes d'erreurs de principe ϵ^1_i et ϵ^2_i sur les graphiques G.V.9 et G.V.10.

Les courbes sont effectivement assez bien stabilisées.

2ème solution :

Au lieu d'avoir un intégrateur qui émet un incrément

$$+ 1 \text{ lorsque } R_j \geq + 1$$

$$0 \text{ lorsque } +1 > R_j > - 1$$

$$- 1 \text{ lorsque } - 1 > R_j$$

On peut imaginer un intégrateur qui émet

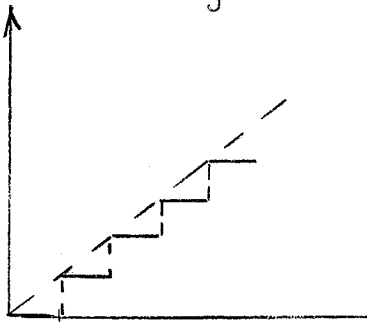
$$+ 1 \text{ lorsque } R_j \geq + 1$$

$$0 \text{ lorsque } + 1 > R_j > 0$$

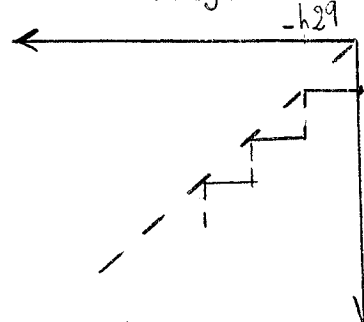
$$- 1 \text{ lorsque } 0 \geq R_j$$

La loi d'arrondi est :

1) Tous les y_j sont > 0

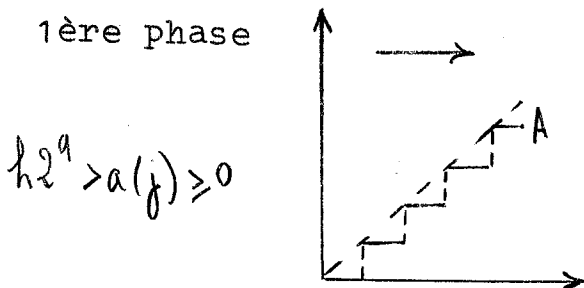


2) tous les y_j sont < 0

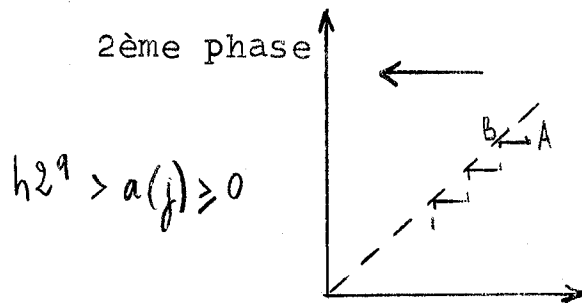


dans le cas où y_j passe d'une valeur positive à une valeur négative, on a :

1ère phase



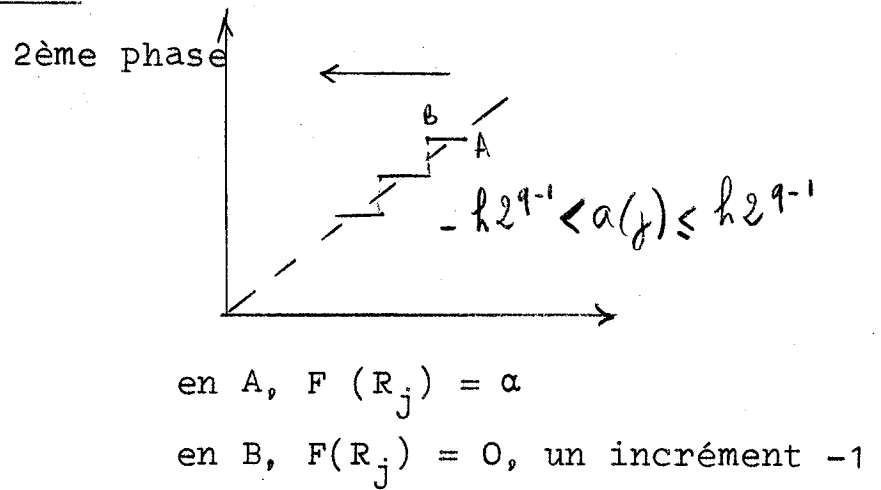
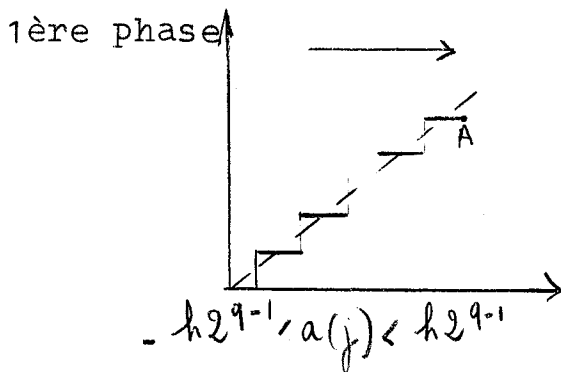
2ème phase



Pour faire varier $a(j)$ entre les bornes $\pm h 2^{q-1}$, il suffit de mettre une valeur $+\frac{1}{2}$ dans le registre R.

$k = +\frac{1}{2}$ quel que soit y_j puisque $F(R_j)$ est toujours positif.

Passage d'un $y_j > 0$ à un $y_j < 0$



est émis.

La loi de variation de l'arrondi est excellente.

CALCUL DES RACINES D'UNE EQUATION
DU SECOND DEGRE

Soit une équation du second degré $a x^2 + b x + c = 0$ ayant des racines réelles, ces racines peuvent être obtenues en recherchant les solutions d'une équation différentielle du type

$$\epsilon y' = a y^2 + b y + c \quad (\epsilon = \pm 1)$$

plus précisément en faisant le calcul

$$y_{i+1} = y_i + \epsilon h (a y_i^2 + b y_i + c) = \mathcal{F}(y_i, \epsilon)$$

Dans le cas où cette récurrence, définie à partir d'une valeur initiale y_0 converge vers la valeur α

(la convergence a lieu sous la condition $\mathcal{F}'(\alpha, \epsilon) < 1$)

On a :

$$\alpha = \mathcal{F}(\alpha) \quad \text{d'où } a \alpha^2 + b \alpha + c = 0$$

α est donc une racine de l'équation du 2^e degré.

VI. A. 1. Représentation de la récurrence

Sur l'ADD, la récurrence est représentée sous la forme

$$y_i = y_{i-1} + h y'_{i-1}$$

$$y_i = y'_{i-1} + h y''_{i-1}$$

Nous avons étudié le cas où $\epsilon = -1$ et nous avons noté

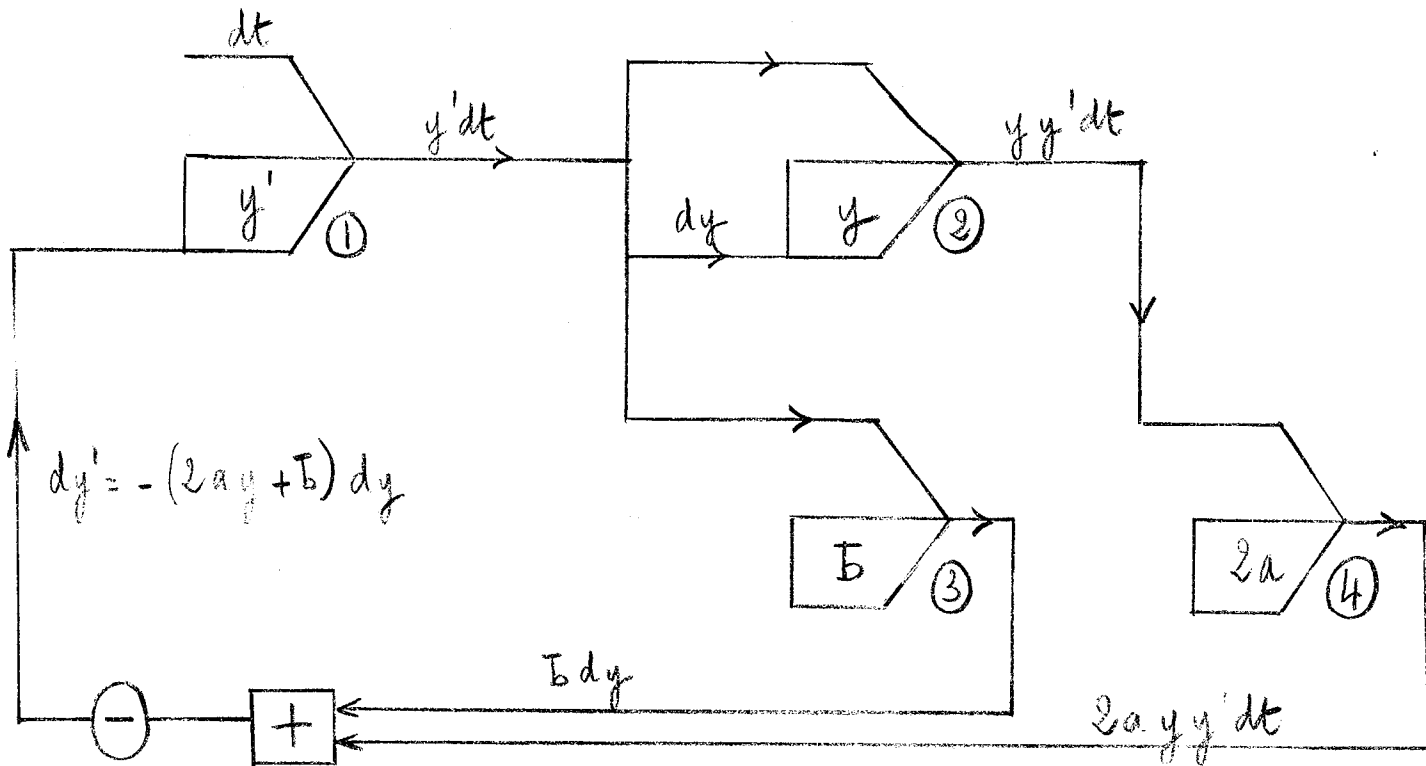
$$y'_{i-1} = y^1_{i-1} \quad \text{et } y''_{i-1} = y^2_{i-1}$$

Le système devient alors

$$y^1_i = y^1_{i-1} - h y^1_{i-1} (2 a y^2_{i-1} + b)$$

$$y^2_i = y^2_{i-1} + h y^1_{i-1}$$

Il est représenté par le schéma suivant



Les valeurs initiales sont $y^1_0 = y'_0 = -(ax^2_0 + bx_0 + c)$

et $y^2_0 = x_0 = y_0$.

VI. A. 2. Relations fondamentales

Les relations fondamentales sont :

pour l'intégrateur (1)
$$Y^1_i = Y^1_{i-1} - h 2^{q_3} [E(R^3_{i-1}) + E(R^4_{i-1})]$$

avec $q_3 = q_2 + q_4$

$$R^1_i = F(R^1_{i-1}) + Y^1_i$$

pour l'intégrateur (2)
$$Y^2_i = Y^2_{i-1} + h 2^{q_1 - q_2} E(R^1_i)$$

$$R^2_i = F(R^2_{i-1}) + E(R^1_i) Y^2_i$$

(l'addition du registre Y^2 au registre R^2 n'a lieu au pas i que si $E(R^1_i) \neq 0$)

pour le multiplieur (3)
$$Y^3_i = Y^3_0 = 2^{-q_3} b$$

$$R^3_i = F(R^3_{i-1}) + E(R^1_i) Y^3_i$$

pour le multiplieur (4)
$$Y^4_i = Y^4_0 = 2^{-q_4 + 1} . a$$

$$R^4_i = F(R^4_{i-1}) + E(R^2_i) Y^4_i$$

Poids des incréments émis

$$E(R^1_i) \text{ a un poids } 2^{q_1} h$$

$$E(R^2_i) \text{ a un poids } 2^{q_1 + q_2} h$$

$$E(R^3_i) \text{ a un poids } 2^{q_1 + q_3} h$$

$$E(R^4_i) \text{ a un poids } 2^{q_1 + q_2 + q_4} h$$

Les facteurs de cadrage q_3 et q_4 sont choisis de telle manière que $E(R^3_i)$ et $E(R^4_i)$ aient même poids.

et :

$$C_{i-1} = \begin{pmatrix} h 2^{q_1} \left[\lambda^1_{i-1} (2a(2^{q_2} Y^2_{i-1}) + b) \right] + 2a 2^{q_2} \lambda^2_{i-1} + 2^{q_3} (\lambda^3_{i-1} + \lambda^4_{i-1}) \\ - h 2^{q_1} \lambda^1_i + h 2^{q_1} (Y^1_i - Y^1_{i-1}) \end{pmatrix}$$

Ainsi :

$$Y_i = \begin{pmatrix} 0 \\ \pi \\ i-1 \end{pmatrix} A_K Y_0 + \sum_{j=2}^i \begin{pmatrix} i-j+1 \\ \pi \\ i-1 \end{pmatrix} A_K C_{i-j} + C_{i-1}$$

Remarque :

Au pas 0, le registre Y^2 est modifié par l'incrément $E(R^2_0)$

Le contenu de ce registre est $(Y^2_0) = Y^2_0 + h 2^{q_1 - q_2} E(R^2_0)$

et $(\epsilon^2_0) = 2^{q_2} (Y^2_0) - y^2_0 = \epsilon^2_0 + h 2^{q_1} Y^1_0 - h 2^{q_1} \lambda^1_0$

Ainsi :

$$\epsilon^1_1 = \epsilon^1_0 (1 - h(2ay^2_0 + b)) - 2ahy^1_0 (\epsilon^2_0 + h 2^{q_1} Y^1_0 - h 2^{q_1} \lambda^1_0) + h 2^{q_1} H_0$$

$$\epsilon^2_1 = \epsilon^2_0 + h \epsilon^1_0 + h 2^{q_1} Y^1_0 - h 2^{q_1} \lambda^1_0 - h 2^{q_1} \lambda^1_1 + h 2^{q_1} (Y^1_1 - Y^1_0)$$

donc :

$$C_0 = \begin{pmatrix} 2^{q_1} h \left[\lambda^1_0 (2a(2^{q_2} Y^2_0) + b) + 2ah(\lambda^1_0 - Y^1_0) + 2a 2^{q_2} \lambda^2_0 + 2^{q_3} (\lambda^3_0 + \lambda^4_0) \right] \\ - h 2^{q_1} (\lambda^1_0 + \lambda^1_1) + h 2^{q_1} \left[(Y^1_1 - Y^1_0) + Y^1_0 \right] \end{pmatrix}$$

Le vecteur C_{i-j} peut être décomposé de la manière suivante :

$$C_{i-j} = h 2^{q_1} \begin{pmatrix} \lambda_{i-j} (2a(2^{q_2} Y^2_{i-j}) + b) \\ 0 \end{pmatrix} - h 2^{q_1} \begin{pmatrix} 0 \\ 1 \\ \lambda_{i-j+1} \end{pmatrix} + h 2a 2^{q_2 + q_1} \begin{pmatrix} \lambda^2_{i-j} \\ 0 \end{pmatrix} + \dots$$

Etudions, par exemple, l'influence des vecteurs de la

forme $2a \cdot 2^{q_2} h \begin{pmatrix} \lambda^2_{i-j} \\ 0 \end{pmatrix}$ sur la valeur Y_i

Ces vecteurs interviennent dans l'expression de Y_i

sous la forme

$$2a \cdot 2^{q_1+q_2} h \sum_{j=2}^i \begin{pmatrix} i-j+1 \\ \pi \\ i-1 \end{pmatrix} A_K \begin{pmatrix} \lambda^2_{i-j} \\ 0 \end{pmatrix}$$

soit :

$$2a \cdot 2^{q_2+q_1} \cdot h \left[\begin{pmatrix} \lambda^2_{i-1} \\ 0 \end{pmatrix} + A_{i-1} \begin{pmatrix} \lambda^2_{i-2} \\ 0 \end{pmatrix} + \dots + A_{i-1} \dots A_1 \begin{pmatrix} \lambda^2_0 \\ 0 \end{pmatrix} \right]$$

Un produit de matrice $A_{i-1} A_{i-2} \dots A_K$ est une matrice

(2,2) dont nous noterons le terme général

$$a^{i,K}_{m,n}$$

Comme $\lambda^2_{i-j} = F(R^2_{i-j}) - F(R^2_{i-j-1})$

et $\lambda^2_0 = F(R^2_0) - k_2$

(k_2 est la valeur mise initialement dans le registre R^2)

On peut écrire :

$$2a \cdot 2^{q_1+q_2} h \left(\begin{aligned} & F(R^2_{i-1}) - F(R^2_{i-2}) + a^{i,i-1}_{11} [F(R^2_{i-2}) - F(R^2_{i-3})] + \dots \\ & \qquad \qquad \qquad + a^{i,1}_{11} (F(R^2_0) - k_2) \\ & \qquad \qquad \qquad a^{i,i-1}_{21} (F(R^2_{i-2}) - F(R^2_{i-3})) + \dots \\ & \qquad \qquad \qquad + a^{i,1}_{21} (F(R^2_0) - k_2) \end{aligned} \right)$$

Soit :

$$2a \ 2^{q_1+q_2} \ h \left(\begin{array}{l} F(R^2_{i-1})+F(R^2_{i-2}) \left[a_{11}^{i,i-1} - 1 \right] + \dots + F(R^2_0) (a_{11}^{i,1} - a_{11}^{i,2}) \\ - k_2 a_{11}^{i,1} \\ \\ F(R^2_{i-2}) a_{21}^{i,i-1} + \dots + F(R^2_0) (a_{21}^{i,1} - a_{21}^{i,2}) \\ - k_2 a_{21}^{i,1} \end{array} \right)$$

ou encore :

$$2a \ 2^{q_2+q_1} \ h \underbrace{\left(\begin{array}{l} F(R^2_{i-1}) - k_2 \\ 0 \end{array} \right)}_{\text{vecteur arrondi}} + 2a \ 2^{q_1+q_2} \ h \underbrace{\overbrace{\left(F(R^2_j) - k_2 \right)}_{(o, i-2)} \left(\begin{array}{l} a_{11}^{i,1} - 1 \\ a_{21}^{i,1} \end{array} \right)}_{\text{vecteur de propagation de l'arrondi}}$$

Ainsi le vecteur de propagation est le vecteur nul si $\overbrace{F(R^2_j)}_{(o, i-2)} - k_2 = 0$

En faisant un raisonnement similaire pour les autres vecteurs composant C_{i-j} et en notant :

$$a_j = 2a \ (2^{q_2} \ Y^2_j) + b$$

$$Y_i = \begin{pmatrix} 0 \\ {}_{i-1} \pi_1 \ A_K \end{pmatrix} Y_0 \quad \left. \vphantom{Y_i} \right\} \begin{array}{l} \rightarrow \\ \epsilon_{po} \ (i) \end{array}$$

$$\begin{aligned}
 & + h 2^{q_1} \begin{pmatrix} (F(R^1_{i-1}) - k_1) \alpha_{i-1} \\ 0 \end{pmatrix} \\
 & + h 2a 2^{q_1+q_2} \begin{pmatrix} F(R^2_{i-1}) - k_2 \\ 0 \end{pmatrix} \\
 & + h 2^{q_1+q_3} \begin{pmatrix} F(R^3_{i-1}) - k_3 \\ 0 \end{pmatrix} \\
 & + h 2^{q_1+q_2+q_4} \begin{pmatrix} (F(R^4_{i-1}) - k_4) \\ 0 \end{pmatrix} \\
 & - h 2^{q_1} \begin{pmatrix} F(R^1_i) - k_1 \\ 0 \end{pmatrix} \\
 & + h 2^{q_1} \begin{pmatrix} 0 \\ Y^1_i - Y^1_{i-1} \end{pmatrix}
 \end{aligned}
 \left. \vphantom{\begin{aligned} & + h 2^{q_1} \begin{pmatrix} (F(R^1_{i-1}) - k_1) \alpha_{i-1} \\ 0 \end{pmatrix} \\ & + h 2a 2^{q_1+q_2} \begin{pmatrix} F(R^2_{i-1}) - k_2 \\ 0 \end{pmatrix} \\ & + h 2^{q_1+q_3} \begin{pmatrix} F(R^3_{i-1}) - k_3 \\ 0 \end{pmatrix} \\ & + h 2^{q_1+q_2+q_4} \begin{pmatrix} (F(R^4_{i-1}) - k_4) \\ 0 \end{pmatrix} } \right\} \begin{matrix} \xrightarrow{\epsilon_a(i)} \\ \\ \\ \\ \\ \end{matrix}$$

$$\left. \vphantom{\begin{pmatrix} 0 \\ Y^1_i - Y^1_{i-1} \end{pmatrix}} \right\} \xrightarrow{\epsilon_r(i)}$$

Nous ne nous attarderons pas à étudier les vecteurs de propagation des arrondis et de retard dont les termes sont difficilement calculables.

Nous avons par contre défini une nouvelle loi de composition des arrondis.

VI A.4 LOIS DE COMPOSITION DES ARRONDIS

D'après l'expression de $\overrightarrow{\epsilon}_a (i)$

$$\begin{aligned} \epsilon^1_a(i) = & h 2^{q_1} \left[F(R^1_{i-1}) - k_1 \right] \alpha_{i-1} + h 2^{q_1+q_2} 2 a \left[(F(R^2_{i-1}) - k_2) \right] \\ & + h 2^{q_1+q_3} \left[F(R^3_{i-1}) - k_3 \right] + h 2^{q_1+q_2+q_4} \left[F(R^4_{i-1}) - k_4 \right] \end{aligned}$$

et :

$$\epsilon^2_a(i) = - h 2^{q_1} \left[F(R^1_i) - k_1 \right]$$

Nous avons vu que les incréments émis par les opérateurs 1,2,3,4 avaient respectivement les poids

$$h 2^{q_1}, h 2^{q_1+q_2}, h 2^{q_1+q_3}, h 2^{q_1+q_2+q_4}$$

Donc :

$$\epsilon^1_a(i) = a_1(i-1) \left[2a (2^{q_2} Y^2_{i-1}) + b \right] + 2a a_2(i-1) + a_3(i-1) + a_4(i-1)$$

$$\epsilon^2_a(i) = - a_1(i)$$

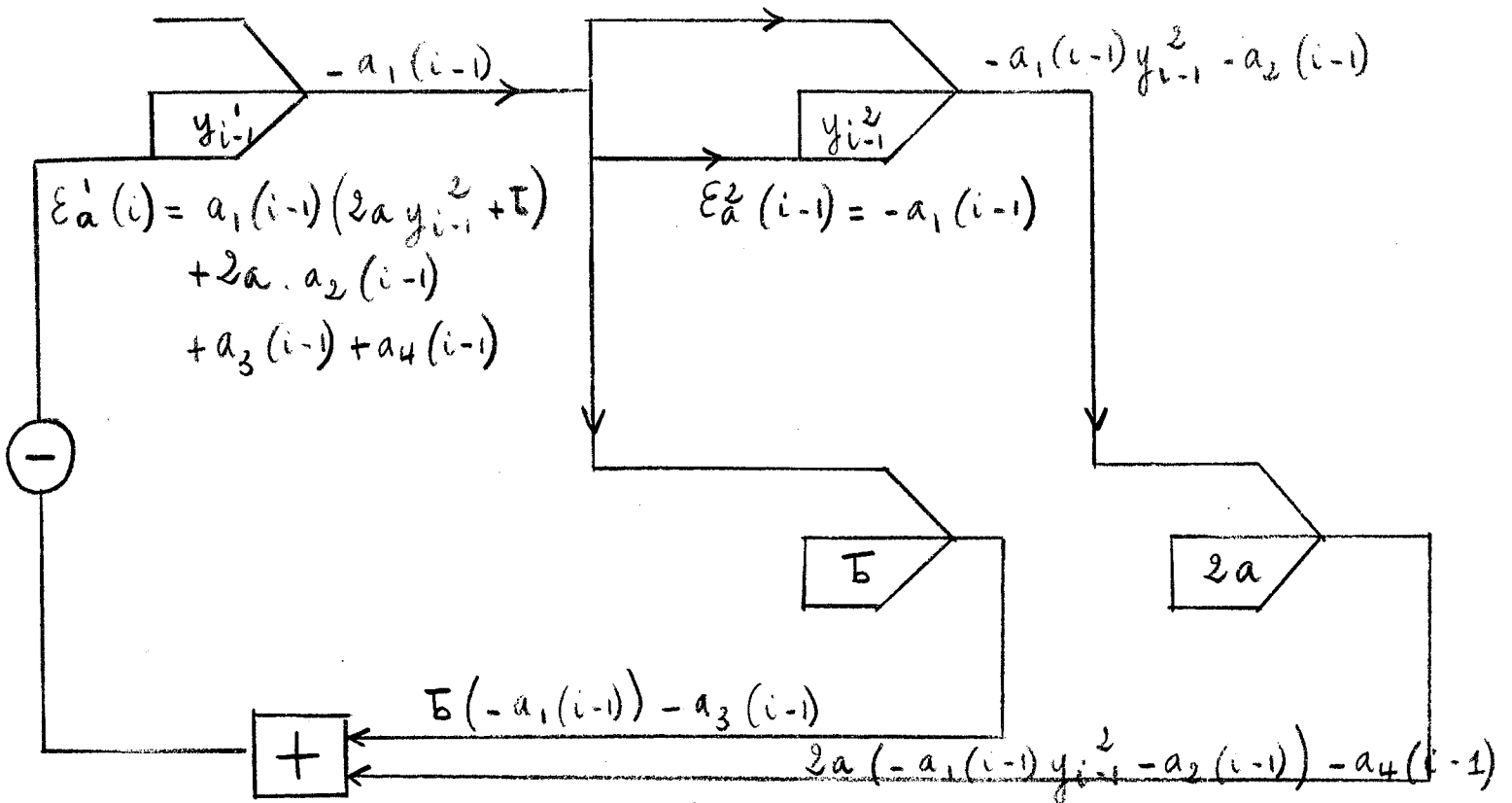
Le terme $a_1(i-1) (2^{q_2} Y^2_{i-1})$ exprime que l'intégrateur

② est alimenté par l'incrément émis par l'intégrateur ①

La loi de composition est la même que pour un multiplicateur mais le terme multiplicatif est variable.

En utilisant les lois de composition déjà définies, on peut résumer le système (1) par le schéma suivant

Au pas $i-1$



VI. B VERIFICATION NUMERIQUE

Nous avons recherché la solution positive α_1 de l'équation du second degré

$$3,25 x^2 + 2,5 x - 1 = 0$$

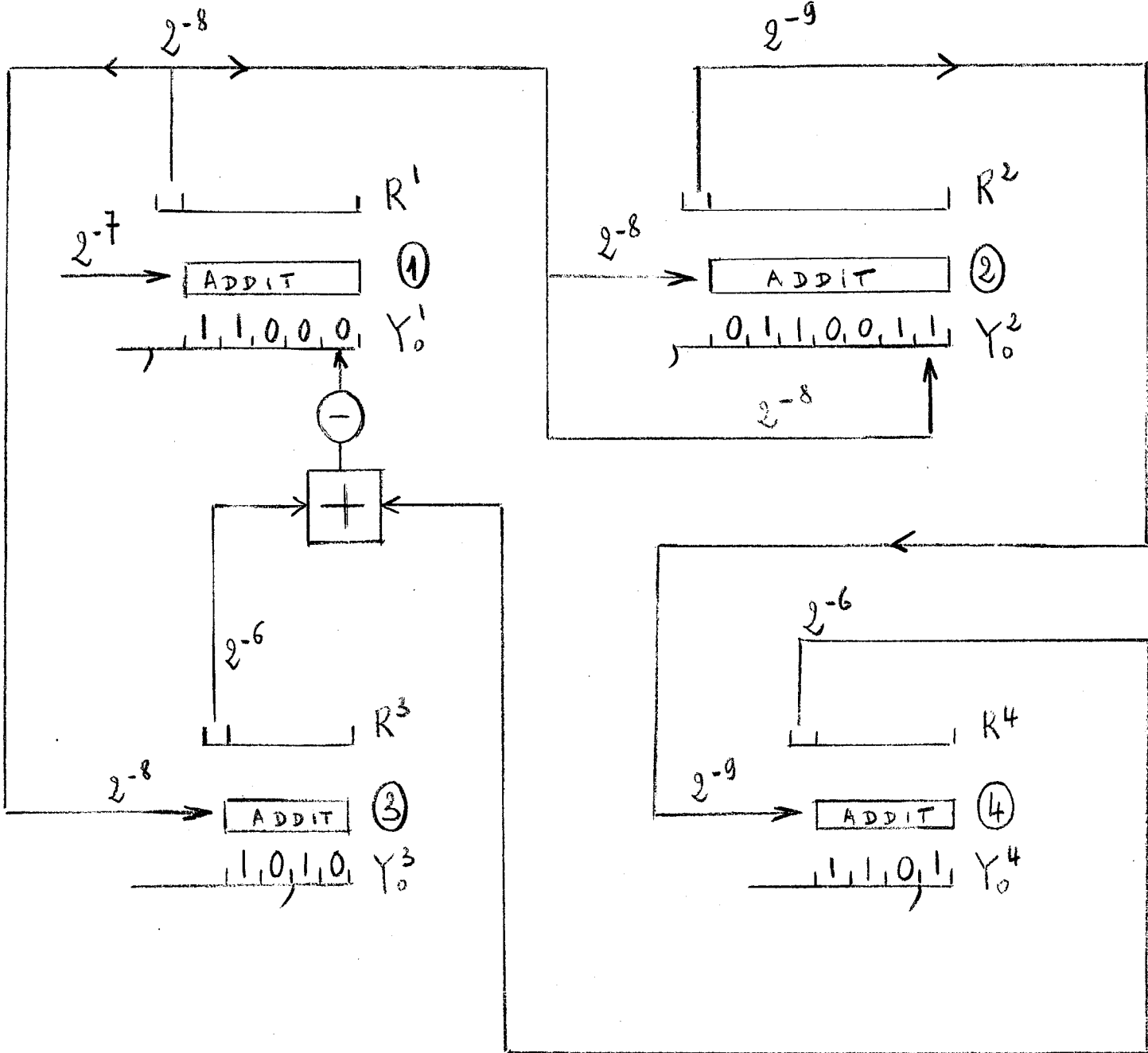
La relation de récurrence à adopter est :

$$y_{i+1} = y_i - h (3,25 y_i^2 + 2,5 y_i - 1)$$

Nous avons pris

$$h = 2^{-7}, y_0^2 = 0, 19921875 \quad y_0^1 = -0,375$$

$$q_1 = q_2 = 1 \quad , \quad q_3 = 2, \quad q_4 = 3$$



VI. B. 1 INTEGRATIONS PAR LA METHODE DES RECTANGLES

Sur le graphique G.VI.1, nous avons tracé la courbe représentative de l'erreur de principe ϵ^2_i , $\frac{1}{2}$ ayant été mis dans les registres R avant le calcul.

Nous avons vu que nous ne pouvons définir théoriquement l'équation de la courbe de propagation ; cette expérience nous prouve donc simplement que l'erreur de principe ne varie pas régulièrement de part et d'autre de l'axe des x.

VI. B. 2 INTEGRATIONS PAR LA METHODE DES TRAPEZES

Nous avons comparé les valeurs données par l'ADD aux solutions y^1_i et y^2_i du système de récurrence.

$$y'_i = y'_{i-1} + h \left[y''_{i-1} + \frac{1}{2} (y''_{i-1} - y''_{i-2}) \right]$$

$$y_i = y_{i-1} + h (y'_i - \frac{1}{2} (y'_i - y'_{i-1}))$$

soit, en adoptant les notations $y^1_i = y'_i$ et $y^2_i = y_i$

$$y^1_i = y^1_{i-1} - h \left[(2ay^2_{i-1} + b) y^1_{i-1} + \frac{1}{2} \left[2a \Delta y^2_{i-1} y^1_{i-1} + (2a y^2_{i-1} + b) \Delta y^1_{i-1} \right] \right]$$

$$y^2_i = y^2_{i-1} + h \left[y^1_i - \frac{1}{2} (y^1_i - y^1_{i-1}) \right]$$

avec $\Delta y^k_{i-1} = y^k_{i-1} - y^k_{i-2}$

Expériences :

Cas ① $K_i = 0$

Le terme de propagation des erreurs d'arrondi n'est pas éliminé. Si l'erreur de principe ϵ^1_i ne dérive pas (graphique G.VI.2

et G. VI. 6), il n'en est pas de même de ϵ^2_i (graphique G.VI.7)

Cas 2 $k_i = \frac{1}{2}$

Le calcul commence au pas 0 pour tous les intégrateurs ; le terme de propagation des erreurs d'arrondi est éliminé.

Théoriquement, on a :

$$Y_i = \begin{pmatrix} \epsilon^1_i \\ \epsilon^2_i \end{pmatrix} = \overrightarrow{\epsilon_a(i)} + \overrightarrow{\epsilon_{pr}(i)}$$

et d'après les lois de composition

$$\epsilon^1_a(i) = a_1(i-1)(2a y^2_{i-1} + b) + 2a \cdot a_2(i-1) + a_3(i-1) + a_4(i-1)$$

soit encore :

$$\left| \epsilon^1_a(i) \right| < 2^{-9} (2 a y^2_{i-1} + b) + 2a \cdot 2^{-10} + 2^{-7} + 2^{-7}$$

$$\left| \epsilon^1_a(i) \right| < 0,03$$

On peut constater sur le graphique G.VI.3 que $\epsilon^1_i < 0,03$;

$\epsilon^1_{pr}(i)$ a donc assez peu d'influence sur ϵ^1_i

Nous noterons que les limites de $\epsilon^1_a(i)$ sont assez larges ; elles supposent que les contenus des registres R^1, R^2, R^3, R^4 prennent la valeur 0 (ou 1) simultanément.

Nous remarquerons cependant que très souvent :

$$\left| \epsilon^1_i \right| = \left| \epsilon^1_a(i) \right| > \text{Max} \left| a_3(i-1) + a_4(i-1) \right| = 0,015625$$

et qu'il est dangereux de ne retenir pour $\epsilon_a^1(i)$ que l'expression

$$\epsilon_a^1(i) = a_3(i-1) + a_4(i-1).$$

On a de même $\epsilon_i^2 = \epsilon_a^2(i) + \epsilon_{pr}^2(i)$
 $= -a_1(i) + \epsilon_{pr}^2(i)$

$$|a_1(i)| < 2^{-9}$$

$\epsilon_{pr}^2(i)$ exprime l'erreur que l'on commet en considérant qu'à la fin du cycle i , Y^2 contient $2^{-q_2} y_i^2$ (en fait il contient $2^{-q_2} y_{i+1}^2$)

On peut voir sur le graphique G.VI.7 que $|\epsilon_i^2| > 2^{-9}$

Il est donc indispensable d'annuler $\epsilon_{pr}^2(i)$.

Cas ③ $K_i = \frac{1}{2}$, $K'_1 = -0,25$

L'intégrateur ① ne fonctionne pas au cycle 0, mais l'intégrateur ② ne peut pas fonctionner non plus (puisqu'il est alimenté par $E(R_j^1)$]

Le seul intérêt de cette méthode est que au pas j , Y^2 contient effectivement $2^{-q_2} y_j^2$.

Cependant Y^1 ne contient plus que $2^{-q_1} y_{j-1}^1$.

le phénomène de retard n'est donc éliminé.

En fait, il est intéressant d'avoir une bonne évaluation de y_j^2 Sur le graphique G.VI.6 ϵ_i^2 ne varie pas strictement dans les limites définies théoriquement ; cependant la solution est meilleure que dans les cas ① et ②.

Sur le graphique G.VI.4 est représenté ϵ^1_i
qui malgré le phénomène de retard, varie dans la bande définie
théoriquement.

Remarque :

Pour le calcul de la racine, il semble indifférent d'utiliser la 2^e ou la 3^e méthode puisque, à la limite, $y^2_{i+1} = y^2_i$.

CONCLUSION GENERALE

Cette étude nous a montré qu'il était indispensable de réduire les erreurs dans le fonctionnement de l'analyseur différentiel digital; la réduction est obtenue facilement en mettant des valeurs centrales dans les registres R des opérateurs et en tenant compte du phénomène de retard.

A N N E X E

SIMULATION D'UN ANALYSEUR DIFFERENTIEL

DIGITAL PAR UN ORDINATEUR

GAMMA E. T.

Le principe de l'analyseur différentiel^{digital} a été rappelé en introduction ; pour toute précision, le lecteur pourra se reporter aux travaux effectués par MM. Vivier, Du Masle, Gaillard et Villiers.

Dans cette annexe sont décrites, d'une part les caractéristiques essentielles de l'appareil simulé, c'est-à-dire le code de représentation des nombres, les registres, les méthodes d'intégrations utilisées, l'organisation générale de l'appareil, d'autre part les procédés de simulation.

Première Partie

CARACTERISTIQUES DE L'ANALYSEUR SIMULE

L'analyseur différentiel digital simulé est :

séquentiel
binaire
à incréments ternaires
à tambour magnétique.

Le type séquentiel de cet appareil et l'utilisation du système de numération binaire conviennent bien à la simulation sur l'ordinateur Gamma E.T.

A - LES REGISTRES

Leur longueur est de 24 positions binaires.

1. Codification des nombres

L'analyseur ne traite que des nombres n tels que

$$- 1 \leq n < 1$$

- Un nombre $n > 0$ est représenté en convention fractionnaire

$$00, \alpha_1 \alpha_2 \dots \alpha_{22}$$

les 2^0 à l'extrême gauche représentent le signe ; α_{22} a le poids minimum

- Un nombre $n < 0$ est représenté par son complément à 4 c'est-à-dire à 100,00 ... 0

$$\text{ex : } a = 00,10100 \dots 0$$

$$[-a] = 11,01100 \dots 0$$

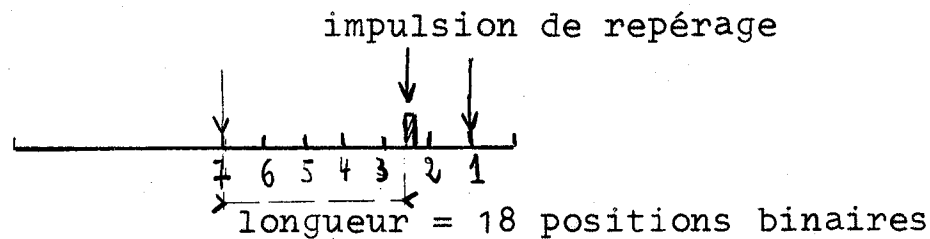
Deux bits à l'extrême gauche représentent donc le signe moins.

particularités des registres Y.

Pour modifier le pas d'intégration, il faut pouvoir ajouter ΔY à Y en position variable.

Pour que les registres Y aient une longueur variable, on place une impulsion de repérage à la droite de la dernière position binaire désirée.

Cette impulsion n'a pas de signification arithmétique. Elle n'est pas traitée par le bloc arithmétique.



B - LES INCREMENTS

Ce sont des quantités ternaires de valeur absolue zéro ou un. Ces quantités sont représentées sur 2^2 avec la convention suivante :

$$0 \longrightarrow 00$$

$$+ 1 \longrightarrow 01$$

$$- 1 \longrightarrow 11$$

La mémoire d'incrément est à accès rapide (mémoires 4, 5, 6, 7)

1) Constitution de ΔY

ΔY est la somme de 7 incréments

$$- 7 \leq \Delta Y \leq + 7$$

Les incréments étant représentés sur 2^2 , il faut procéder à un décodage de manière à ce que ΔY soit un nombre binaire de longueur variable.

α) On définit $b = \sum_{i=0}^7 I Y_i$ écrit sur un τ binaire et complément vrai (complément à 2)

Les quantités terniaries deviennent $0 \longrightarrow 0000$

$+1 \longrightarrow 0001$

$-1 \longrightarrow 1111$

β) b est calculé sur un τ qui contient 8 préalablement à tout calcul.

Lorsque sont ajoutés les incréments écrits sur 4θ et que sont abandonnés tous les débordements, on définit un compteur qui évolue de + 15 à +1

Le tableau de variation du compteur est le suivant

b	COMPTEUR								
	décimal	binaire							
		8	4	2 1					
		code machine							
+ 7	+ 15	1	1	1	1	0	1	1	1
+ 6	+ 14	1	1	1	0	0	1	1	0
+ 5	+ 13	1	1	0	1	0	1	0	1
+ 4	+ 12	1	1	0	0	0	1	0	0
+ 3	+ 11	1	0	1	1	0	0	1	1
+ 2	+ 10	1	0	1	0	0	0	1	0
+ 1	+ 9	1	0	0	1	0	0	0	1
0	+ 8	1	0	0	0	0	0	0	0
- 1	+ 7	0	1	1	1	1	1	1	1
- 2	+ 6	0	1	1	0	1	1	1	0
- 3	+ 5	0	1	0	1	1	1	0	1
- 4	+ 4	0	1	0	0	1	1	0	0
- 5	+ 3	0	0	1	1	1	0	1	1
- 6	+ 2	0	0	1	0	1	0	1	0
- 7	+ 1	0	0	0	1	1	0	0	1

un débordement positif est repérable par la combinaison 01 à gauche de la virgule.

β) Débordement négatif

Soit $[-a] = 11,0110 \dots 0$

$$\begin{array}{r} \text{faire } [-a] - \frac{1}{2} \quad 11,0110 \\ \quad \quad \quad \quad \quad \quad 11,1000 \\ \hline \quad \quad \quad \quad \quad \quad 1\boxed{10},1110 \end{array}$$

Après rangement, la représentation du signe de a (c'est-à-dire 00) et de $[-a]$ (c'est-à-dire 11) est reconstituée.

C - METHODES D'INTEGRATION

1. Intégration par la méthode des rectangles

$$z_{i+1} - z_i = y_i \cdot h$$

2. Intégration par la méthode des trapèzes

Deux formules :

$$z_{i+2} - z_{i+1} = \left\{ y_{i+1} + \frac{1}{2} (y_{i+1} - y_i) \right\} h \quad (a)$$

$$z_{i+1} - z_i = \left\{ y_{i+1} - \frac{1}{2} (y_{i+1} - y_i) \right\} h \quad (b)$$

Une lettre de fonction β , incluse dans le programme, aiguille sur l'une ou l'autre formule.

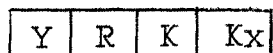
D - ORGANISATION GENERALE

L'appareil simulé possède 40 intégrateurs, 40 multiplieurs et 1 comparateur.

Il faut deux registres pour représenter un intégrateur (registres Y et R), de même pour un multiplieur (registres K et Kx)

Le comparateur réalise une fonction logique.

Les registres sont placés sur le tambour magnétique. A un intégrateur est associé un multiplieur.



Le programme et son rangement sur le tambour :

Dans l'ADD à simuler le programme se présente comme suit.

P_1	1 5		1 5	1 5	1 5	1 5	1 5	1 5			
P_2	2 6		2 6	2 6	2 6	2 6	2 6	2 6			
P_3	3 7		3 7	3 7	3 7	3 7	3 7	3 7			
P_4	4 S		4 S	4 α	4 β	4	4	4			
	IY_1	IY_2	IY_3	IY_4	IY_5	IY_6	IY_7	I_x	I_u	I_z	I_{Kx}

7654321 représente le numéro en binaire de l'intégrateur ou du multiplieur dont on recherche l'incrément, ou dont on veut ranger l'incrément.

S indique le signe à affecter à l'incrément IY_i au moment de la constitution de ΔY

Ex : IY_i (01) si S = 1 prendre (11)

α et β sont des lettres de fonction :

α note le choix entre l'intégrateur et le comparateur

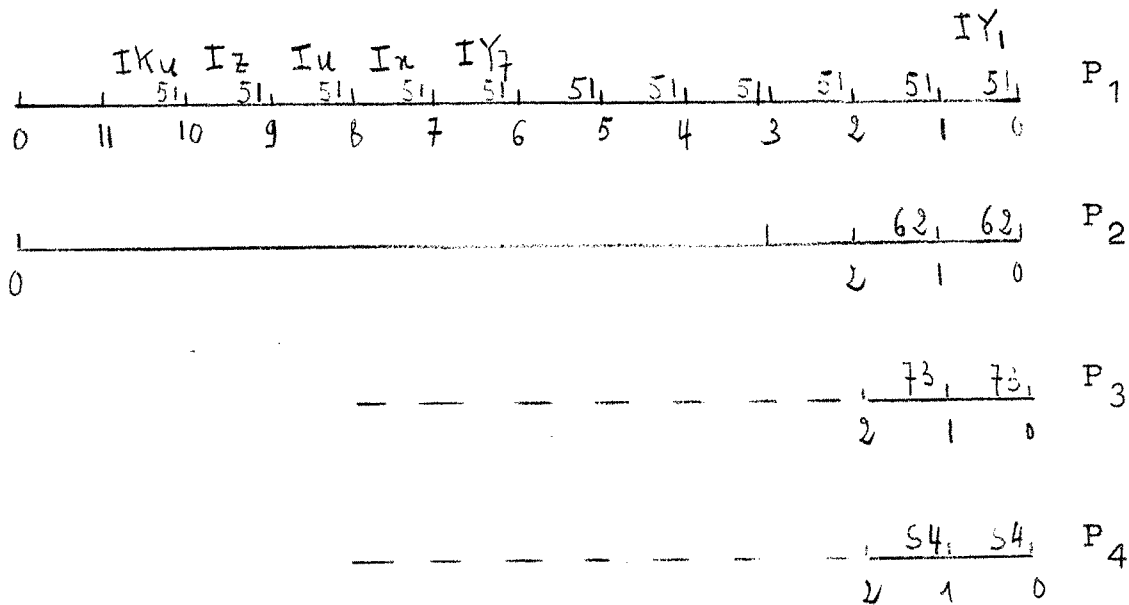
β note le choix entre formules a et b des trapèzes.

Le programme de l'analyseur différentiel digital est normalement inscrit sur un ruban.

De manière à pouvoir éventuellement exploiter le même ruban en simulation (transcription préalable ruban → cartes), le programme a été implanté comme suit sur le tambour de l'ordinateur.

- Le programme nécessaire au traitement d'un cycle mineur (c'est-à-dire d'un intégrateur et d'un multiplieur) est implanté sur une division d'un bloc de tambour.

P_1, P_2, P_3, P_4 sont contenues chacune dans une mémoire



- pendant un cycle mineur

le programme est appelé et décodé

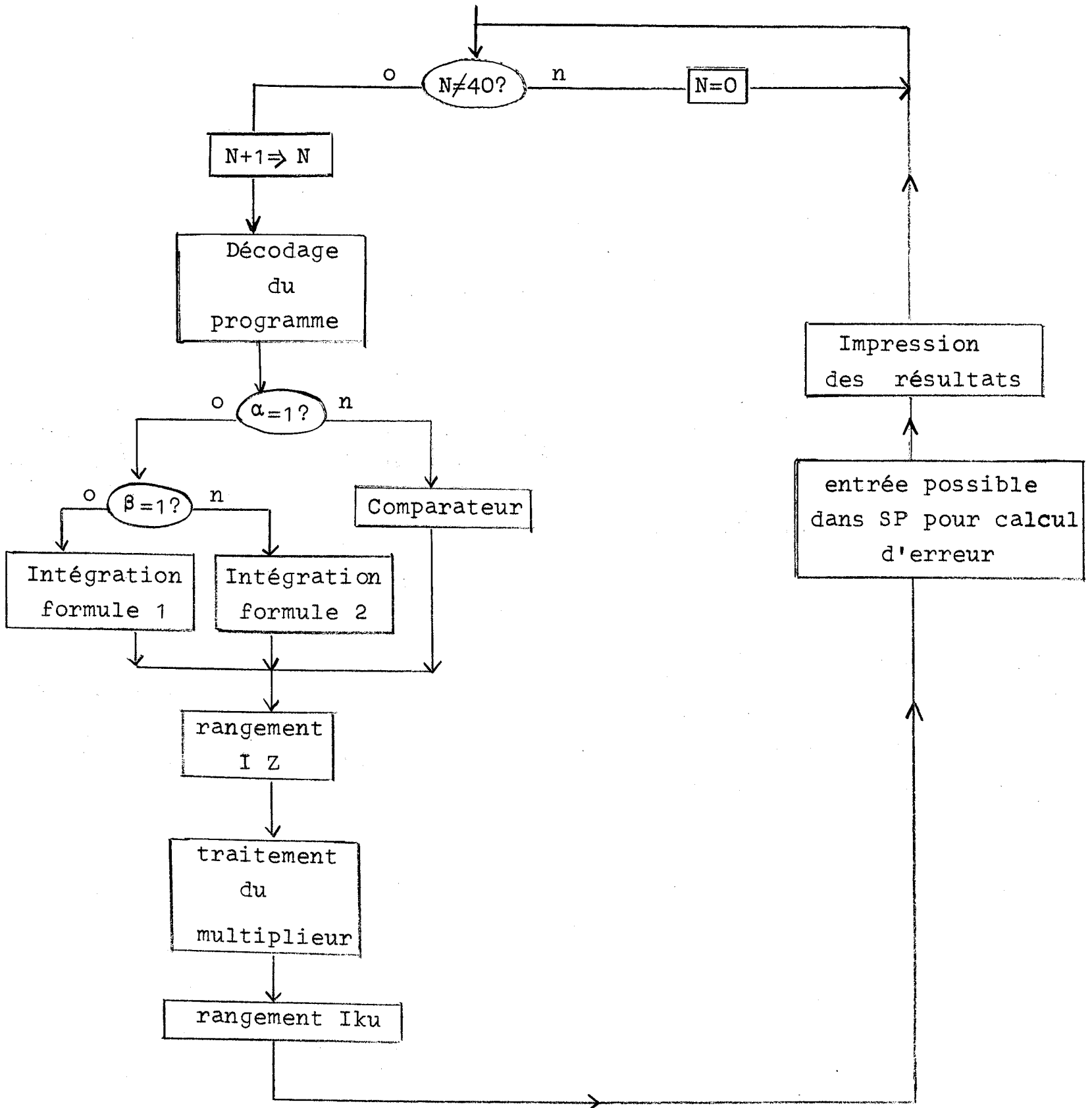
les registres sont traités

les incréments de sortie sont rangés.

Deuxième partie

DESCRIPTION DE LA SIMULATION

ORGANIGRAMME GENERAL



A - L'INTEGRATEUR

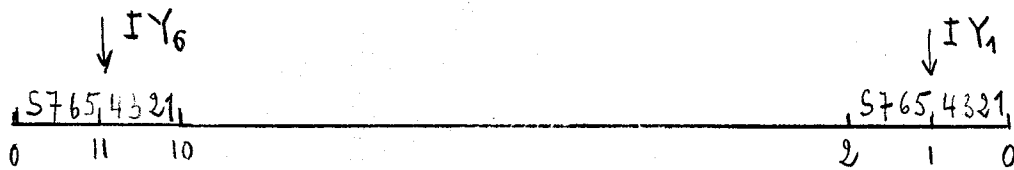
Les informations nécessaires au traitement d'un intégrateur sont placés en octobre 4 au début du cycle mineur.

1. Regroupement du programme en M2

L'adresse d'un incrément a 8 positions binaires

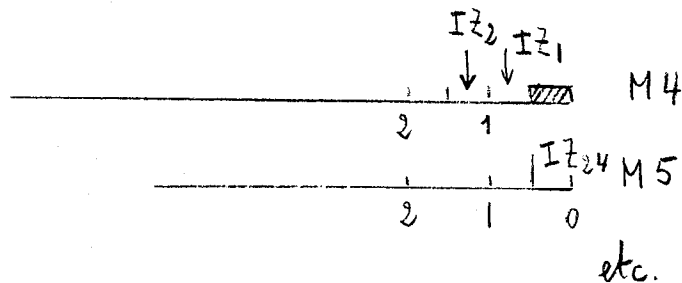
Un premier regroupement est effectué pour les 6 adresses des incréments IY_i (i de 1 à 6)

(suite d'intersections logiques et décalages)



Rappel de la structure de la mémoire d'incrément

Les incréments sont rangés en M 4, 5, 6 ou 7

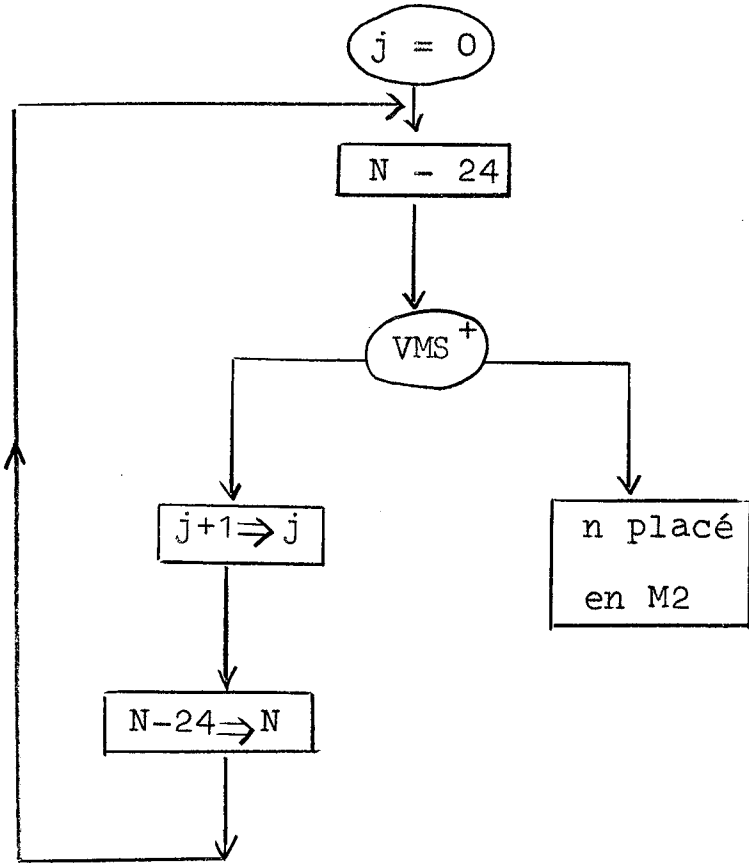


Les adresses des incréments doivent donc être décodées [par ex : l'incrément d'adresse 24 se trouve en M5 de 0 à 1 (partie 1)]

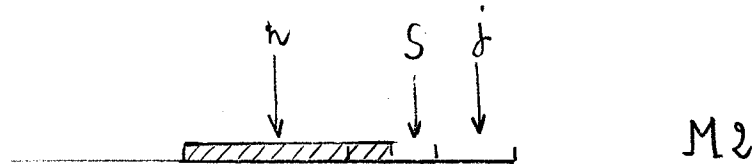
2. Décodage des adresses

Soit N le nombre 7654321 en décimal

Le calcul suivant détermine la position n de l'incrément dans la mémoire j + 4



En fin de décodage on obtient en M2



n peut être pair ou impair ; les incréments des opérateurs à numéro impair sont placés à gauche dans un τ de la mémoire d'incrément.

On remarque alors que n se décompose en 2 parties α et β



avec $0 \leq \beta \leq 11$

$\alpha = 8$ ou 0

β repère la position du τ dans la mémoire ($4+j$)

α la position à gauche ou à droite de l'incrément dans le τ .

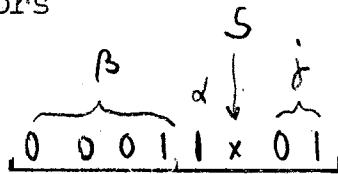
Ex : Soit $IZ_{27} = IY_1$

27 s'écrit en binaire 0011011

$$27 - 24 = 3 > 0 \longrightarrow i = + 1$$

$$3 - 24 > 0 \longrightarrow n = 3 \quad \text{soit } 00011$$

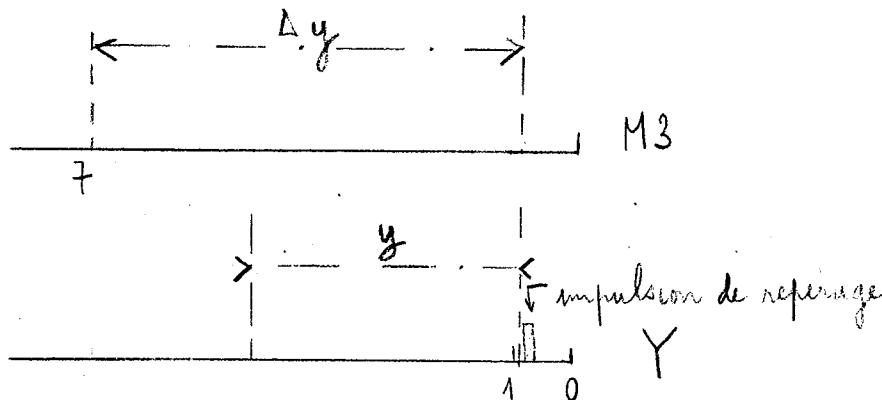
en M2, on obtient alors



Un décodage identique est effectué pour la seconde partie du programme (de IY_7 à IKu)

3. Calcul de $y + \Delta y$

Une impulsion à l'extrême droite de y repère le nombre de positions binaires de y effectivement utilisées.



y est cadré à la position 1 avant l'addition de Δy . Après cette addition, y est remis dans sa position initiale.

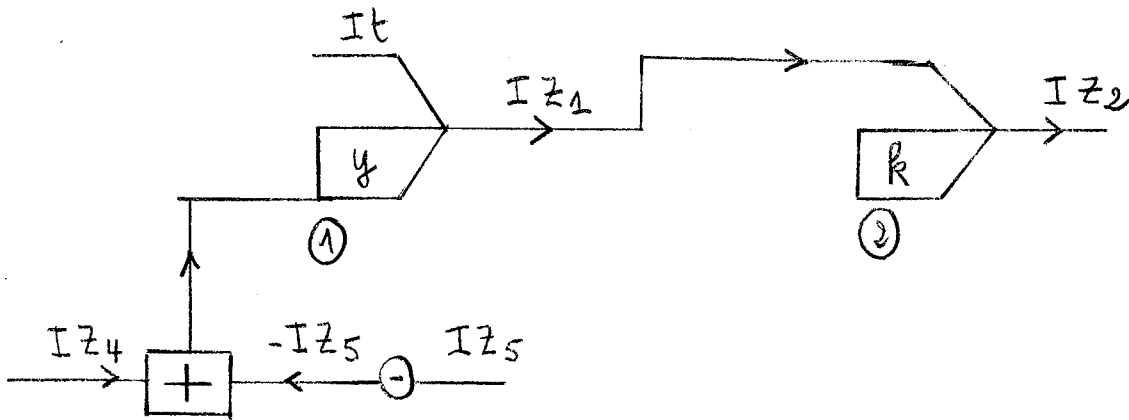
En cas de dépassement de capacité ($y \geq 1$ par ex) la machine imprime le numéro de l'opérateur dont le registre a été dépassé et s'arrête.

4. Intégration par la méthode des trapèzes.

Un test sur β aiguille sur la formule d'avance ou bien de retard.

Le calcul de $\frac{1}{2} \Delta y$ se fait par décalage d'une position binaire à droite.

EXEMPLE D'UTILISATION

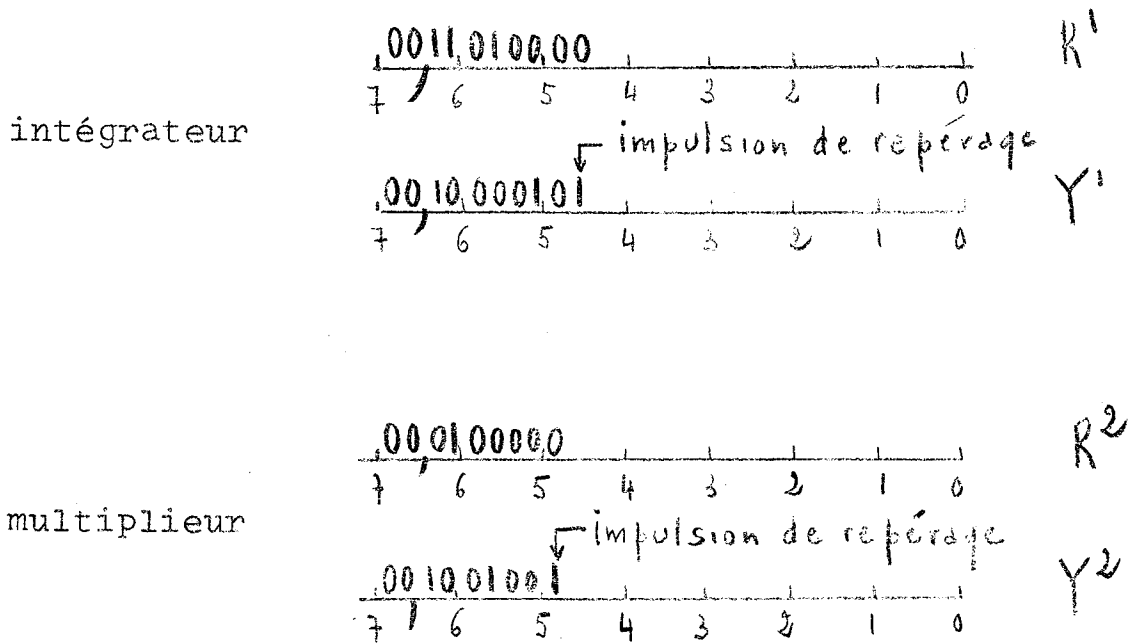


soit à traiter l'intégrateur 1 et le multiplieur 2

Programme

0	0	1	0					0	1	1	0	1	0	0	0
0	0	0	0					1	0	0	0	0	0	1	0
1	0	1	0					1	0	0	0	0	0	0	0
0	0	0	1					0	1	0	0	0	█	0	█
IY ₁		IY ₂						Ix		Iu		IZ		Iku	

Registres



- 107 -

BIBLIOGRAPHIE

- VIVIER : Etude des calculateurs incrémentiels
- DU MASLE : Réalisation de l'organe de calcul d'un analyseur différentiel digital.
(Thèse de 3^e cycle--Mathématiques appliquées GRENOBLE 1961)
- GAILLAID : Réalisation d'une mémoire d'incrément ternaires
(Thèse de 3^e cycle--Servomécanismes GRENOBLE 1961)
- VILLIERS : Contribution à la réalisation d'un analyseur différentiel digital.
Thèse de 3^e cycle servomécanismes GRENOBLE 1963)

Sur les erreurs :

- DANTY : Contribution à l'étude d'un DDA (Analyseur différentiel numérique)
Thèse de 3^e cycle Faculté de TOULOUSE 1960

MILLER, K,S and MURRAY, F.J

"A Mathematical Basis for an Error analysis of Differential Analysers".

MIT Journal of mathematics and Physics, Vol 32.

July October 1953 pp. 136-163

NELSON DON J.

DDA Error Analysis using sampled data techniques
AFIPS 1962 (p. 365 à 373)

Q.C TURTLE

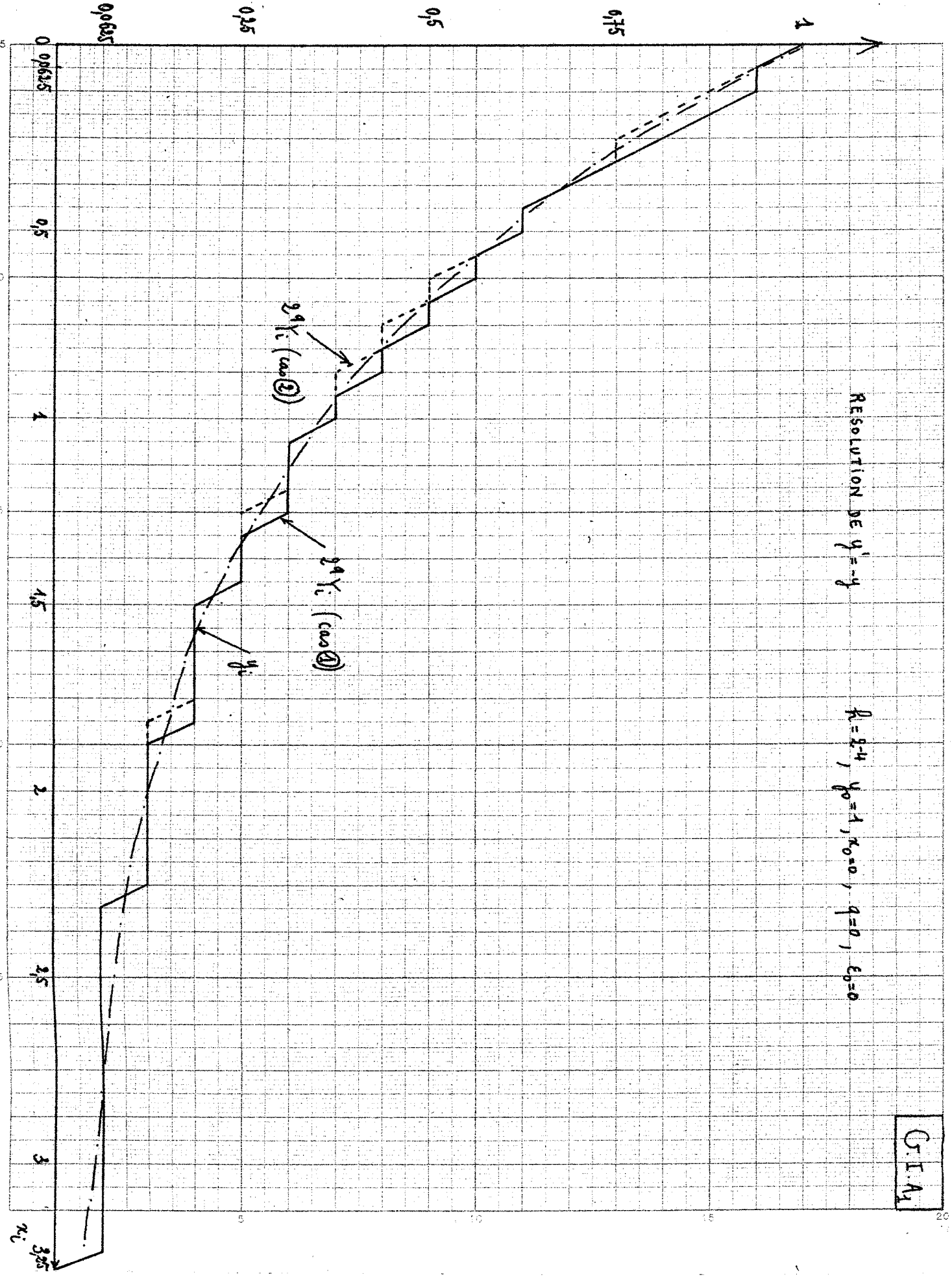
Incremental computer error analysis

IEEE Transactions on communication and electronics

G.I.A.

RESOLUTION DE $y' = -y$

$\mu = 2.4, y_0 = 1, x_0 = 0, q = 0, E_0 = 0$



G.I.A.

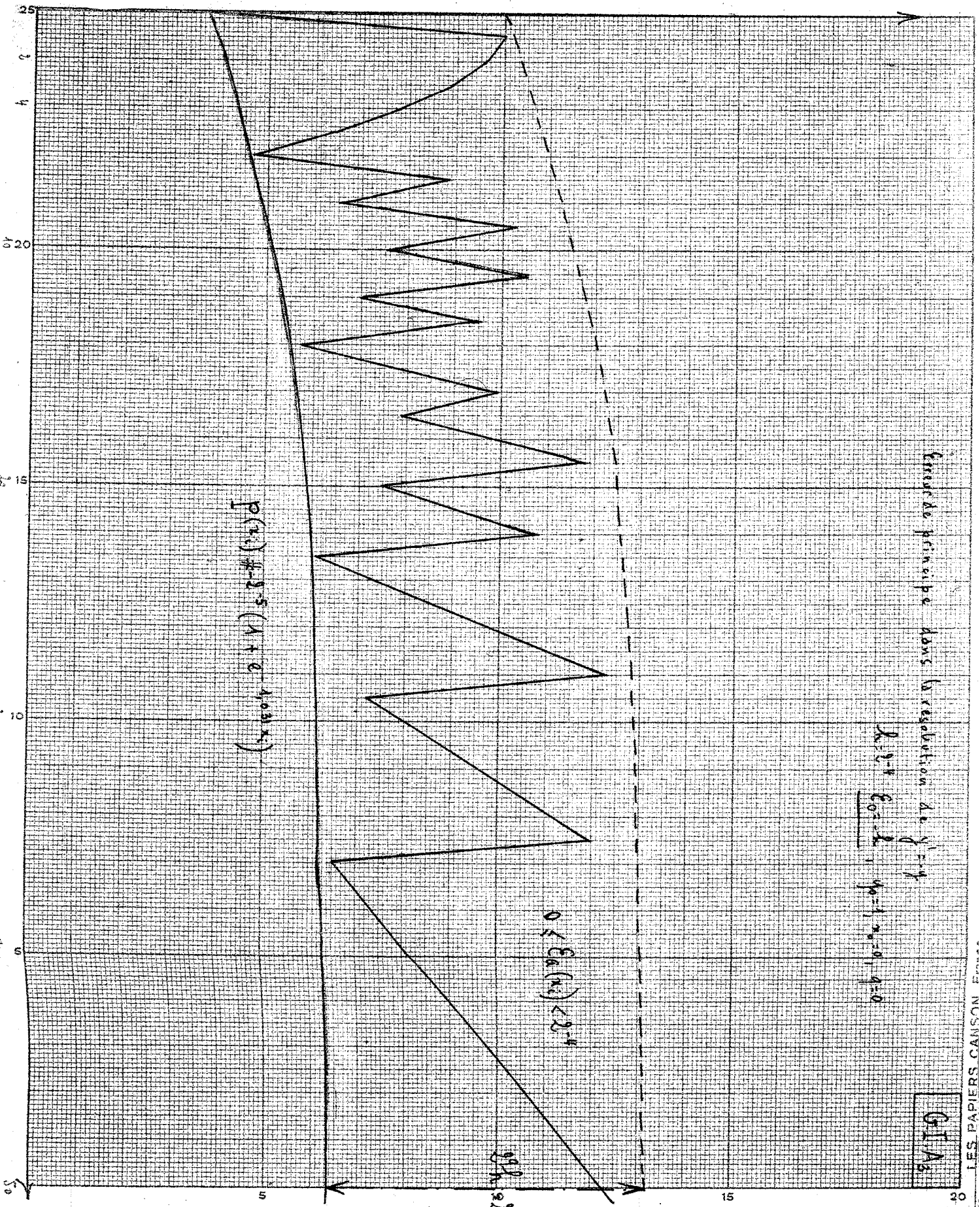
Erreur de principe dans la résolution de $y'' = y$

$y'' = y$
 $y_0 = 1, y_1 = 0, y_2 = 0$

$0 < E_0(n_0) < 2^{-n}$

99
92-4

$P(x) = x^2 - 5(x + 2 + 4000x)$



(1)

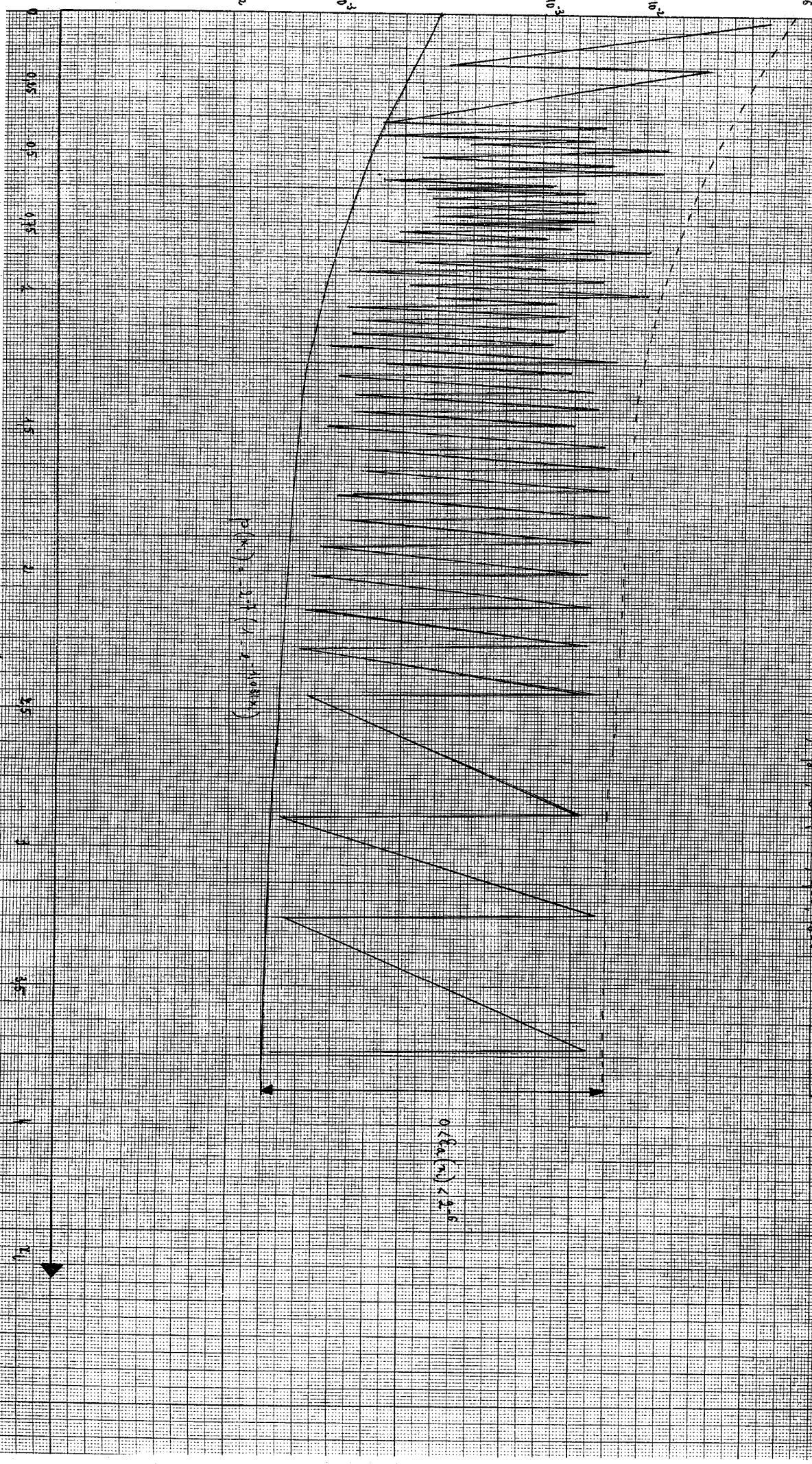
Ensemble average for ensemble resolution $\Delta t = 10^{-6}$ s

Time, t (ns)

G.I.M.

$0 < t < 2.5$

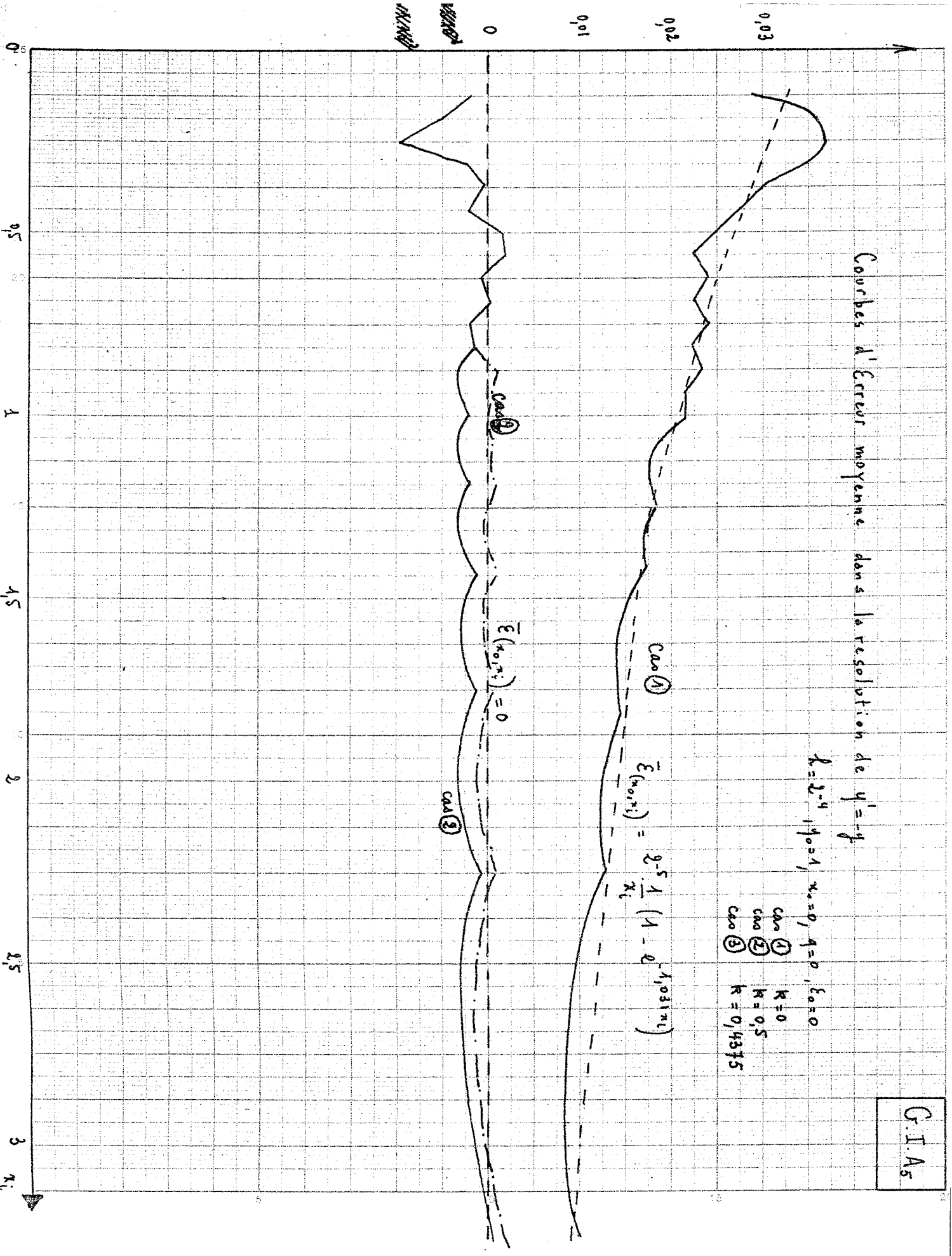
$P(t) = 2.7 \times 10^{-10}$



Courbes d'Erreur moyenne dans la resolution de $y' = -y$

$k = 2^{-4}, \mu_0 = 1, x_0 = 0, q = 0, \epsilon_0 = 0$

- CAA ① $R = 0$
- CAA ② $R = 0,5$
- CAA ③ $R = 0,4375$



CAA ①

$\bar{E}(x_0, \xi_i) = 2^{-5} \frac{1}{\xi_i} (1 - 2^{-4,031 \pi \xi_i})$

$\bar{E}(x_0, \xi_i) = 0$

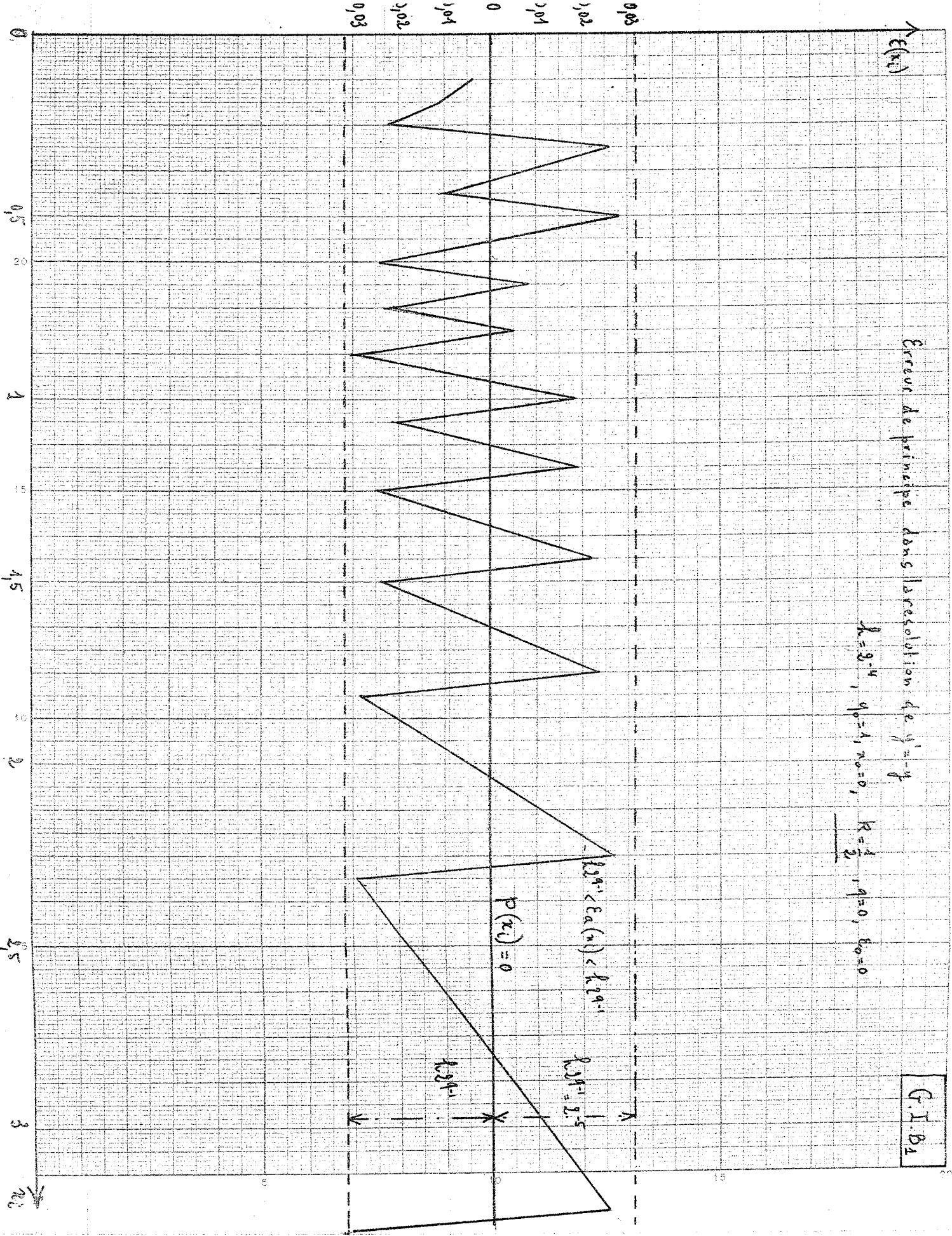
CAA ③

0 0,5 1 1,5 2 2,5 3 ξ_i

0 0,001 0,002 0,003

Erreur de principe dans la résolution de $y'' = y$

$\lambda = 2^{-14}$, $q_0 = 1$, $n_0 = 0$, $R = \frac{1}{2}$, $q = 0$, $\delta_0 = 0$

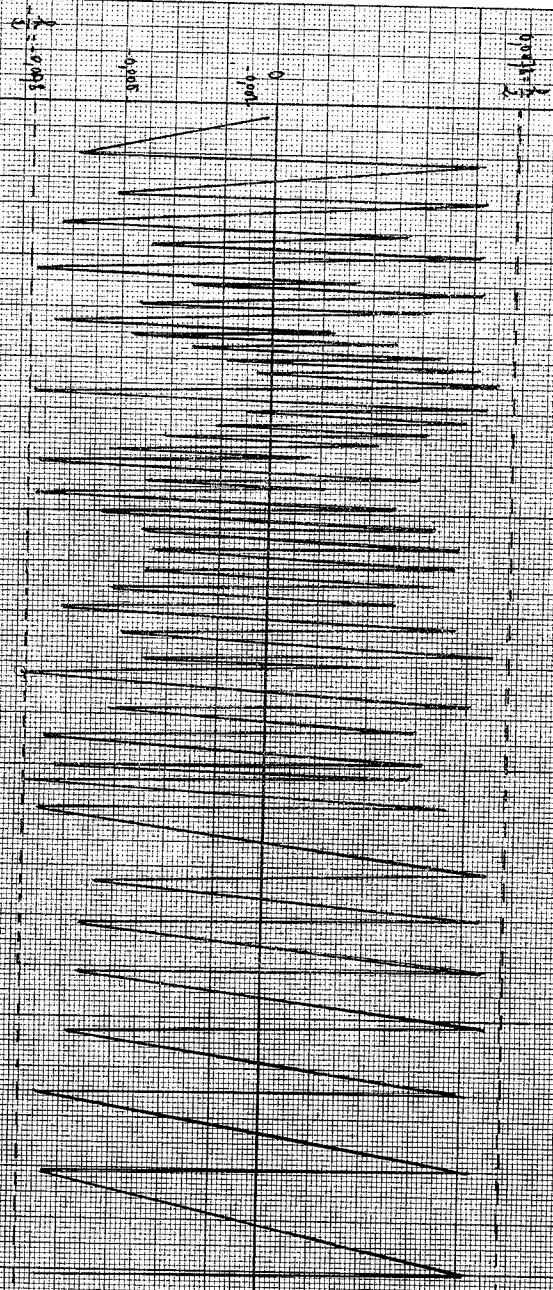


$\epsilon(x)$

Rechercher les principes dans la résolution de $y'' + p(x)y' + q(x)y = r(x)$

Les 10 cas : $K_1, K_2, P_1, P_2, E_1, E_2$

G.I.B.



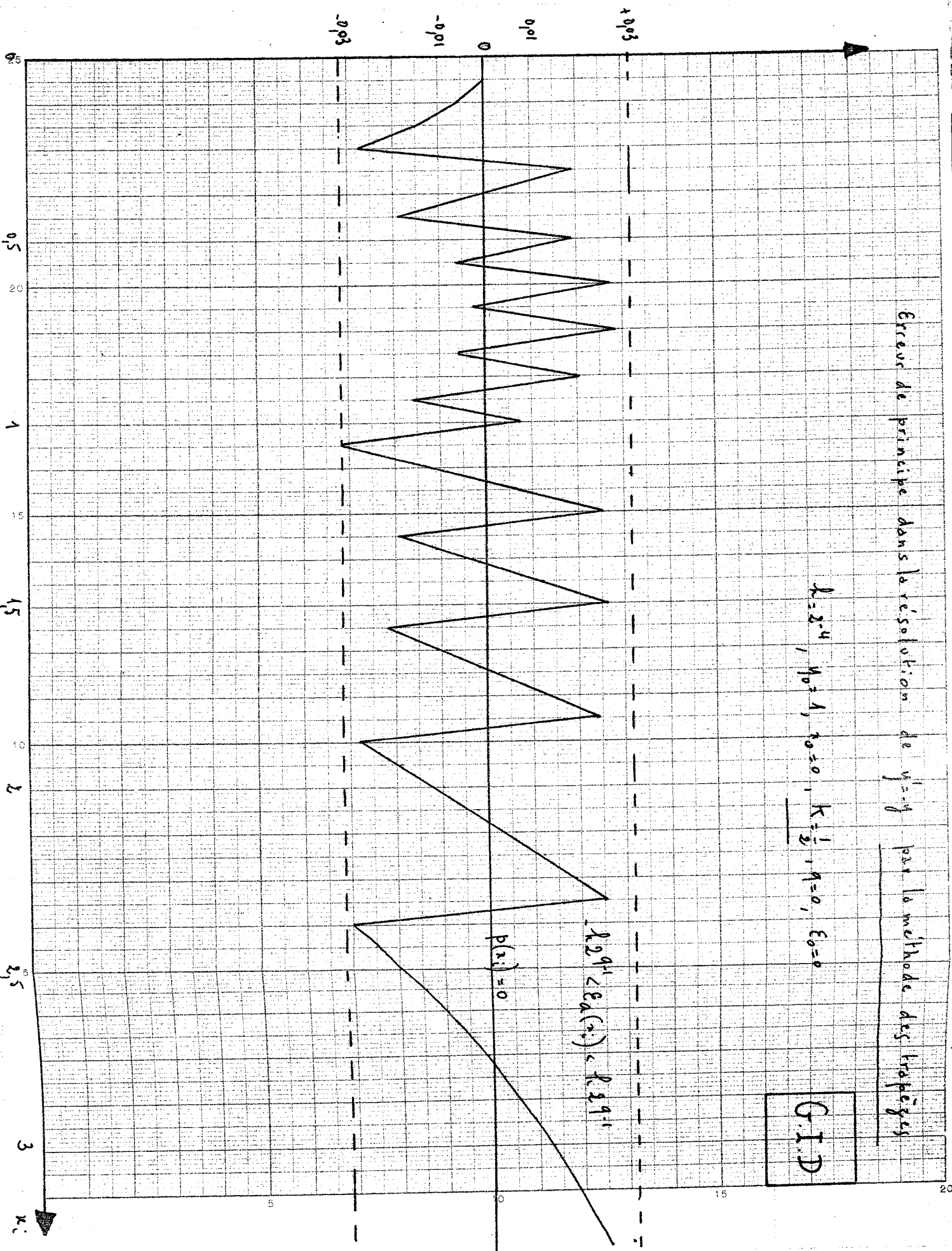
$$y'' + p(x)y' + q(x)y = r(x)$$
$$p(x) = 0$$

0 0.25 0.5 1 1.5 2 2.5 3 x

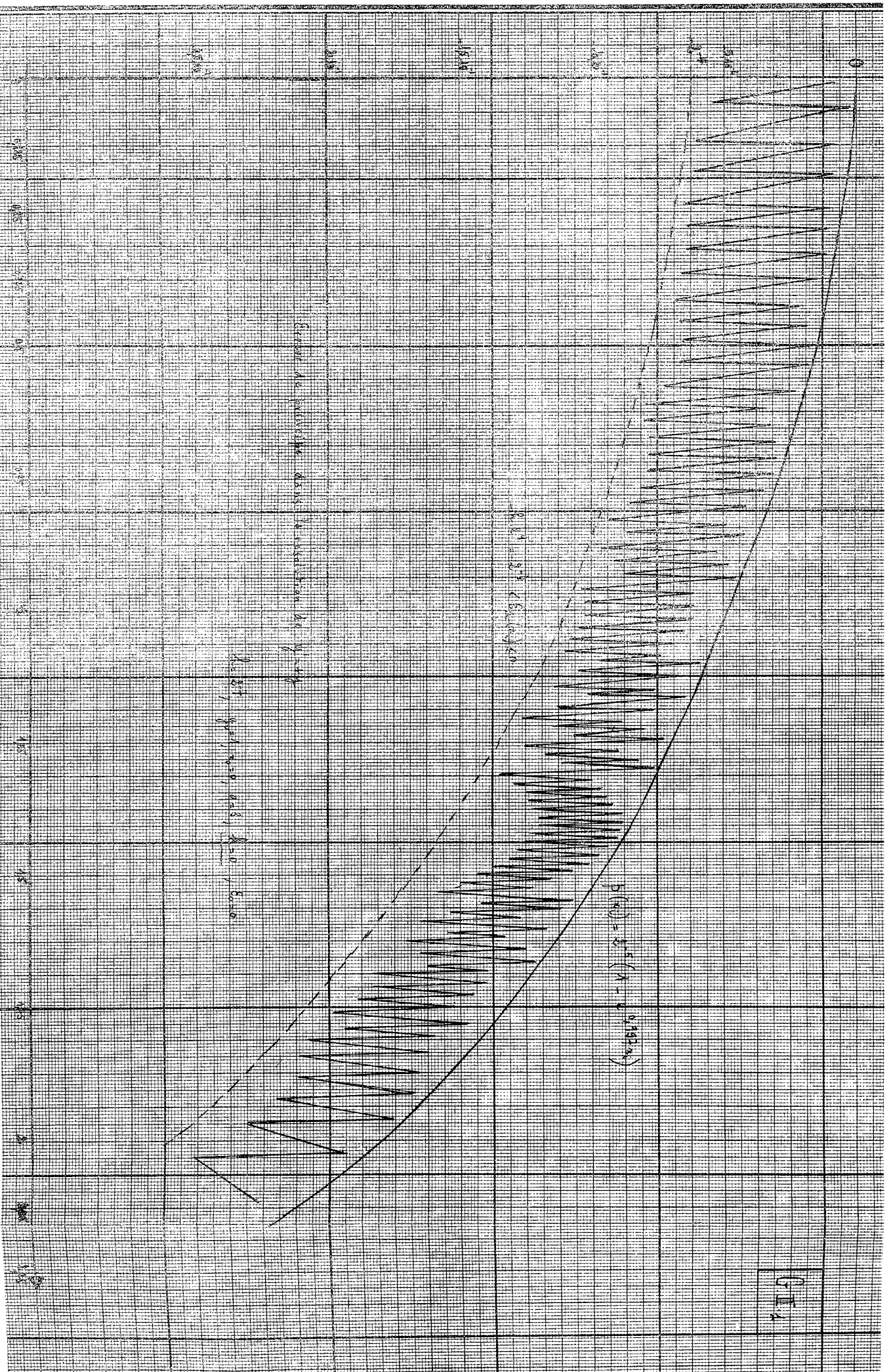
Erreur de principe dans la résolution de $y' = y$ par la méthode des trapèzes

$$h = 2^{-4}, \quad y_0 = 1, \quad y_0 = 0, \quad K = \frac{1}{2}, \quad q = 0, \quad \epsilon_0 = 0$$

G.I.D



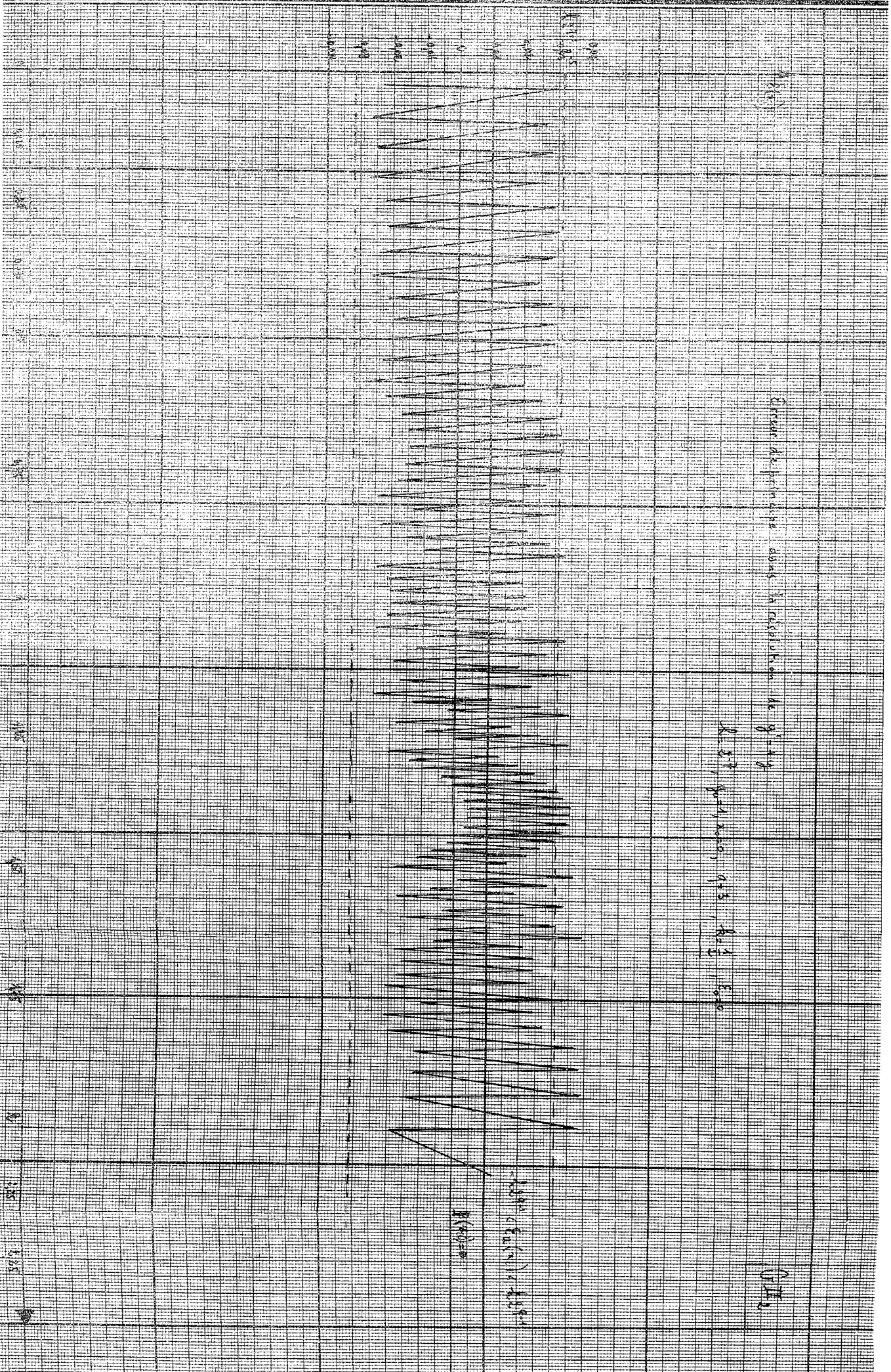
0.1



$D(t) = 1.5 \cdot (1 - 0.1 \cdot \text{abs}(2 \cdot t - 10))$

Graph der periodischen Abszissenfunktion $D(t)$

Amplitude: 1.5, 1.0, 0.5, 0.0, 0.5, 1.0, 1.5



Etalon de fréquence dans la résolution de pi-ny

2.33, 4.10, 9.3, 14.1, 19.0

G.H.

1.99 (0.1) / 2.48

1.99

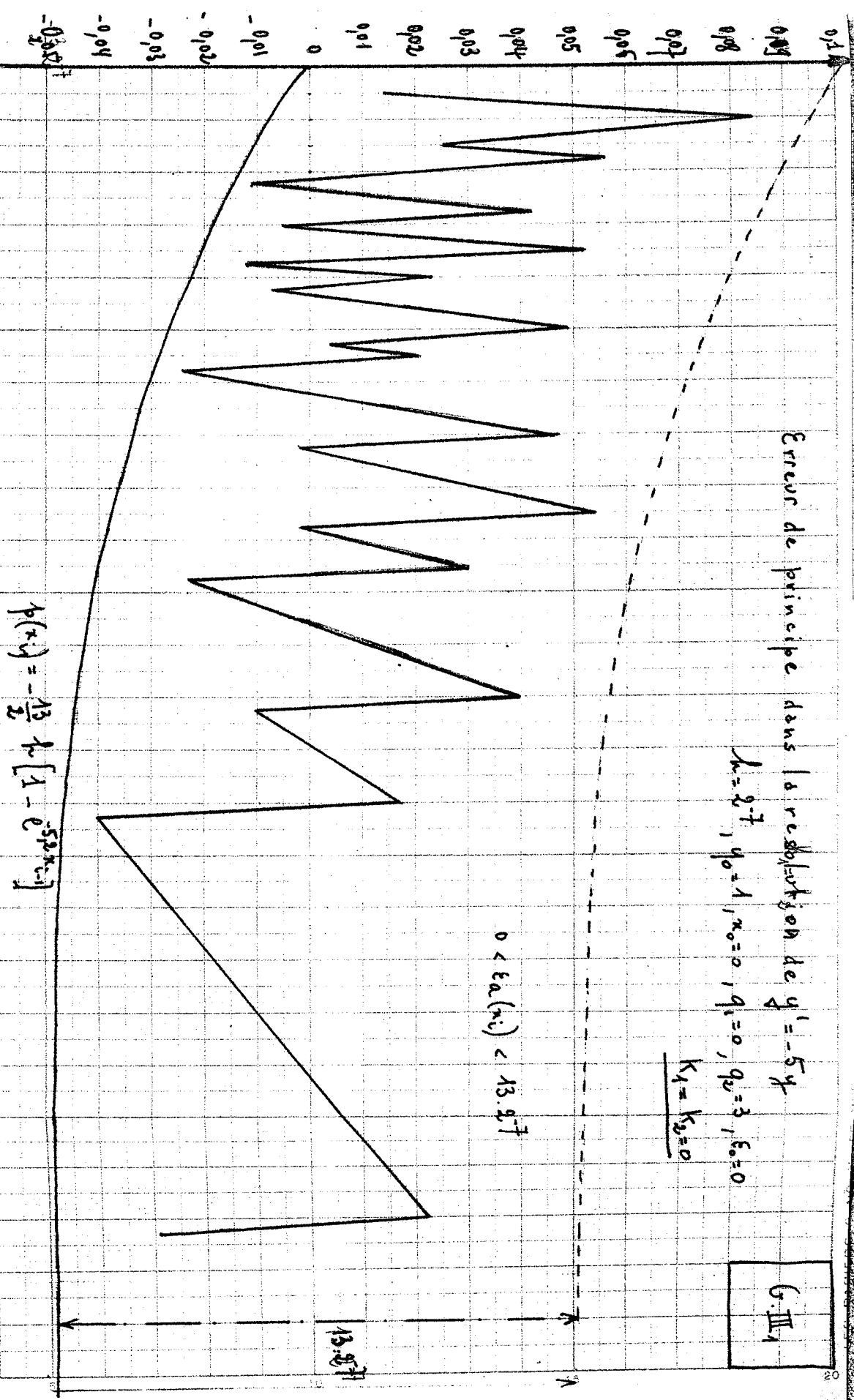
Erreur de principe dans la résolution de $y' = -5y$

$\Delta x = 2.7, y_0 = 1, x_0 = 0, q_1 = 0, q_2 = 3, \epsilon_0 = 0$

$K_1 = K_2 = 0$

$0 < \epsilon_a(x_i) < 13.27$

G-III



$p(x) = -\frac{13}{2} \ln[1 - e^{-5.27x}]$

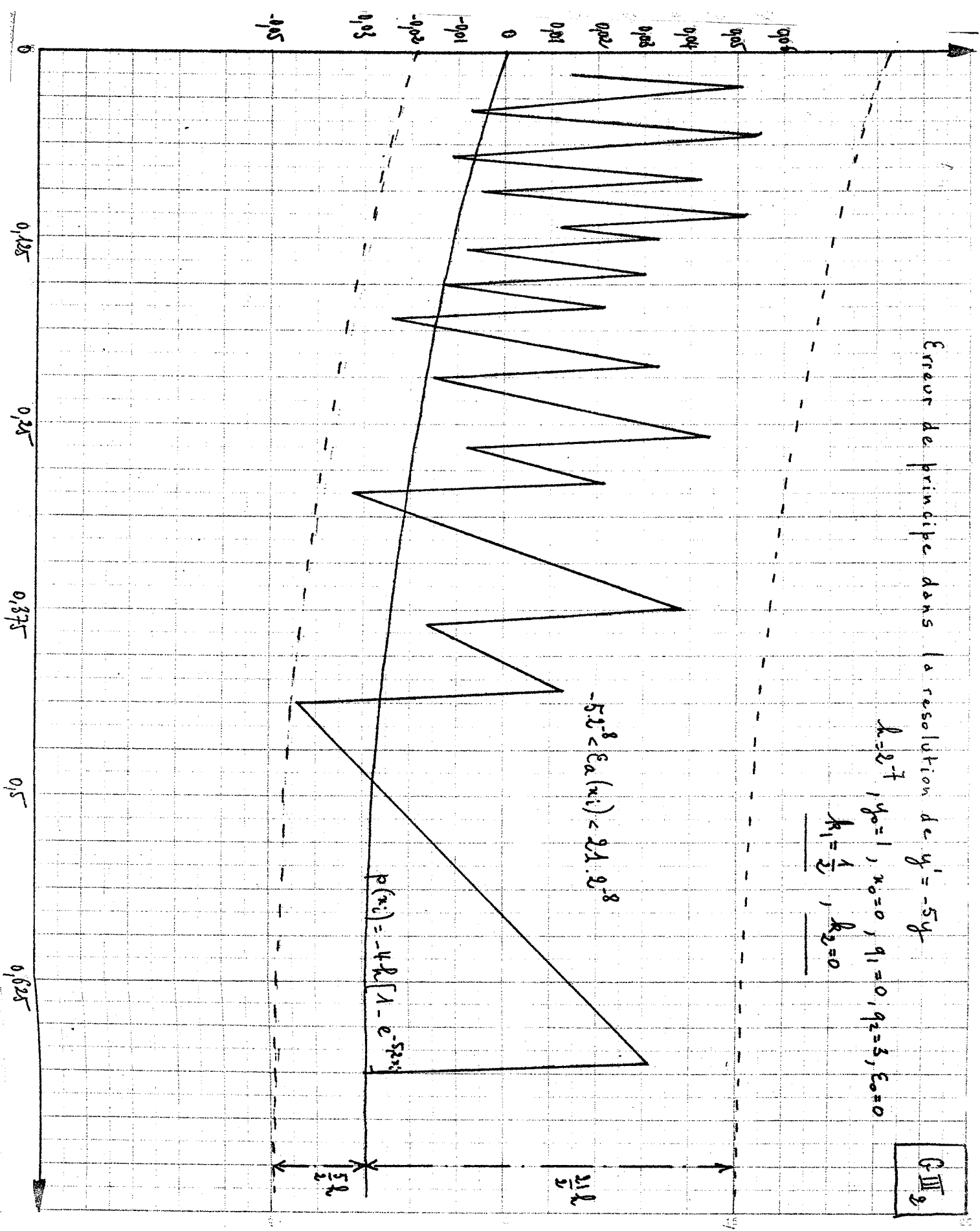
0.125 0.125 0.125 0.125 0.125 0.125

Erreur de principe dans la résolution de $y' = -5y$

$h = 2^{-7}$, $y_0 = 1$, $x_0 = 0$, $q_1 = 0$, $q_2 = 5$, $\epsilon_0 = 0$

$R_1 = \frac{1}{2}$, $R_2 = 0$

C III₂

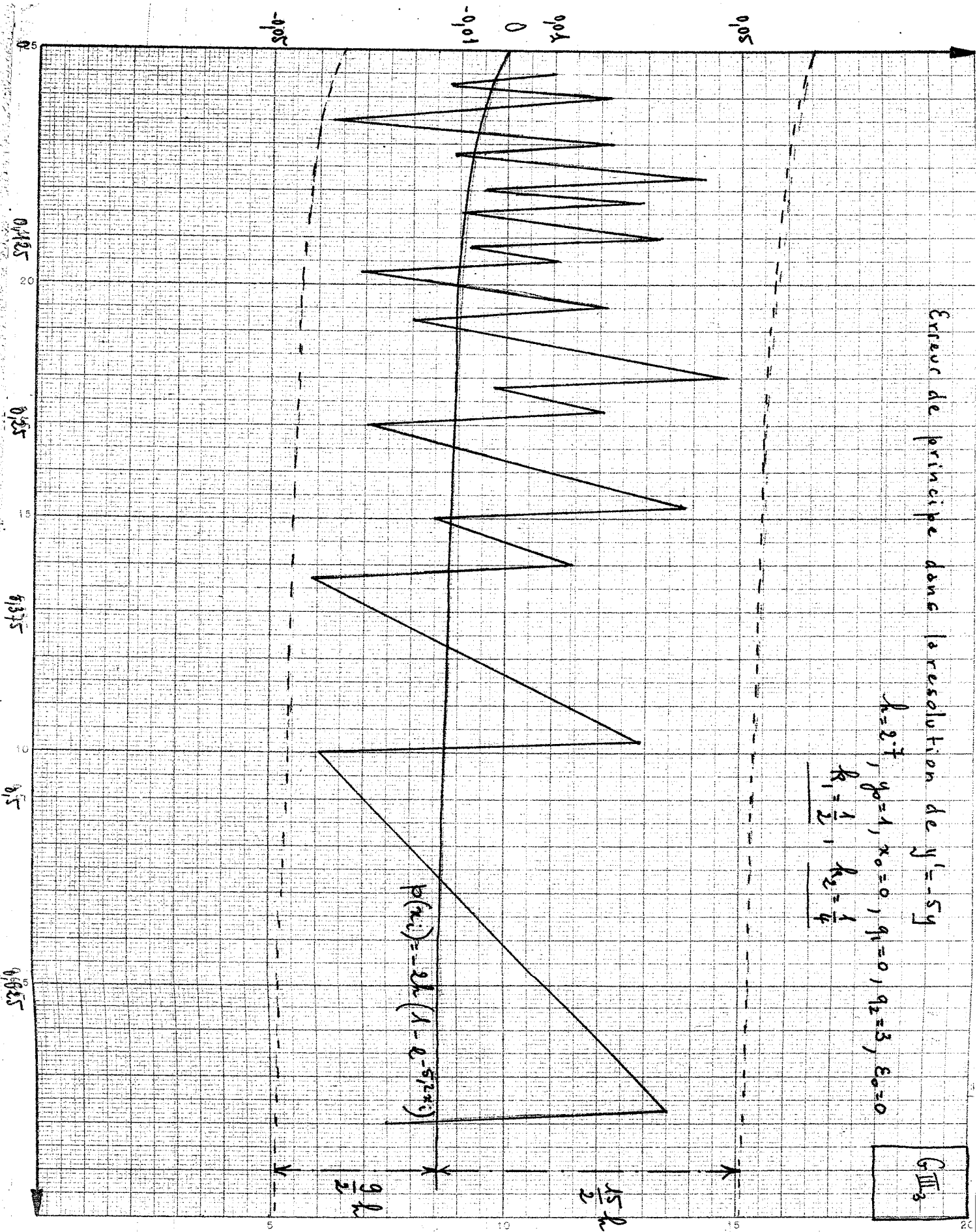


Erreur de principe dans la resolution de $y' = -5y$

$h = 2^{-7}, y_0 = 1, x_0 = 0, q_1 = 0, q_2 = 3, \epsilon_0 = 0$

$k_1 = \frac{1}{2}, k_2 = \frac{1}{4}$

GIII₃



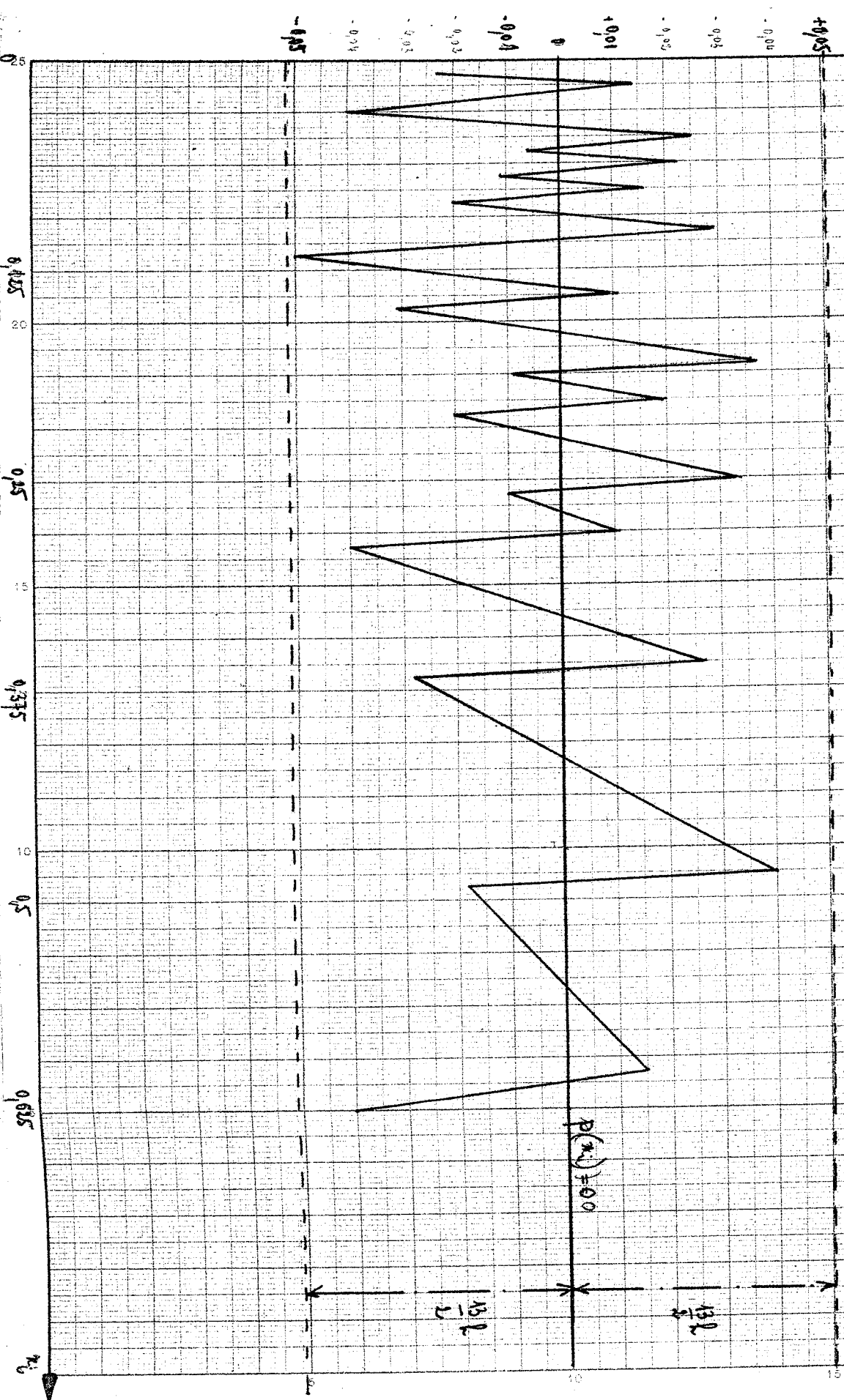
$p(x_i) = -2h(1 - e^{-5(2^i x_i)})$

$\frac{15}{2}h$
 $\frac{9}{2}h$

Erreur de principe dans la resolution de $y' = -5y$

$k = 2.7, \quad \psi_0 = 1, \quad \alpha_0 = 0, \quad q_1 = 0, \quad q_2 = 3, \quad \delta_0 = 0.$

$\underline{R_1 = R_2 = \frac{1}{2}}$



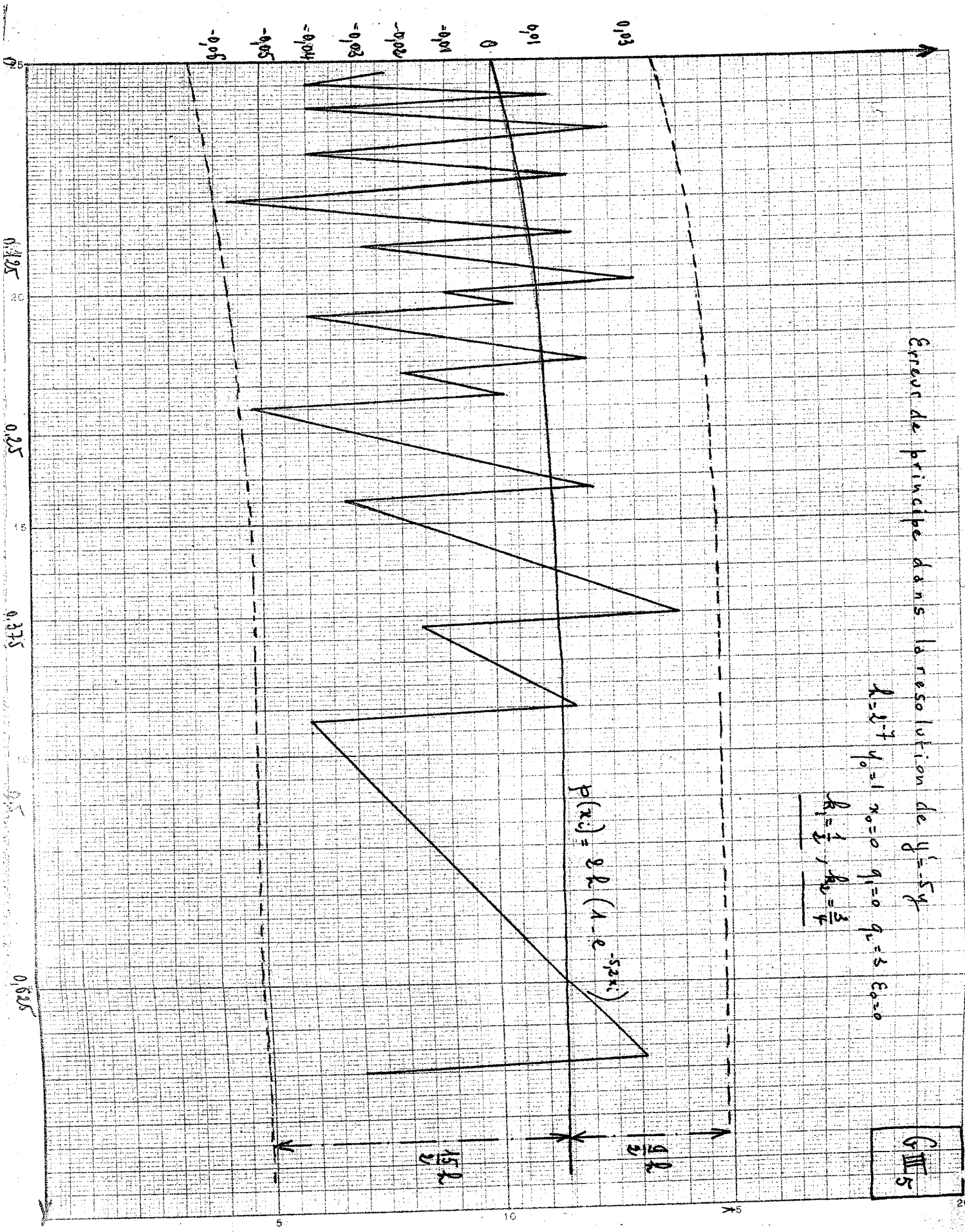
GIII

Erreur de principe dans la résolution de $y' = -5y$

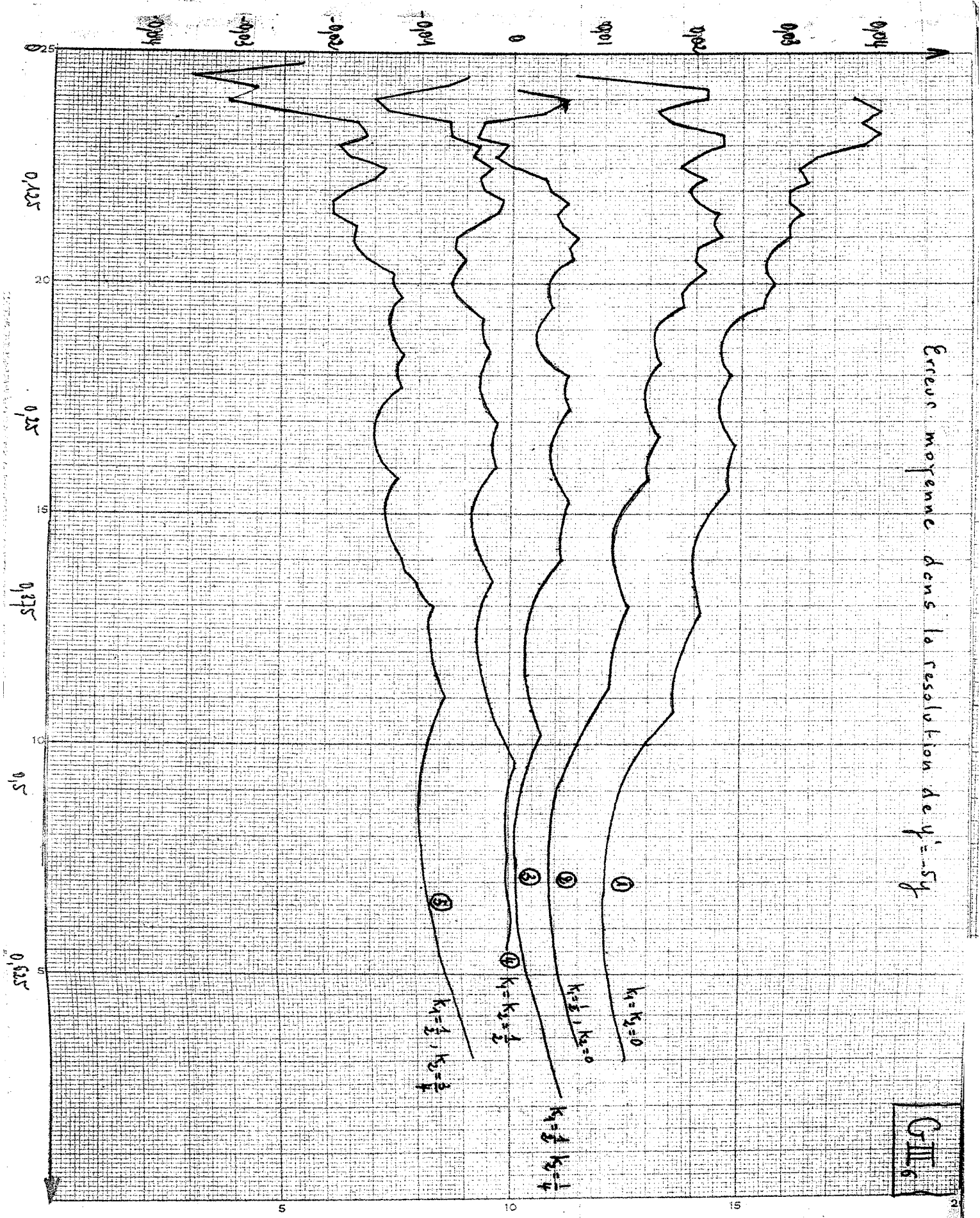
$\lambda = 2.7 \quad y_0 = 1 \quad x_0 = 0 \quad q_1 = 0 \quad q_2 = 3 \quad \epsilon_0 = 0$

$R_1 = \frac{1}{2}, R_2 = \frac{3}{4}$

G III 5



Erreur moyenne dans la resolution de $y' = -5y$



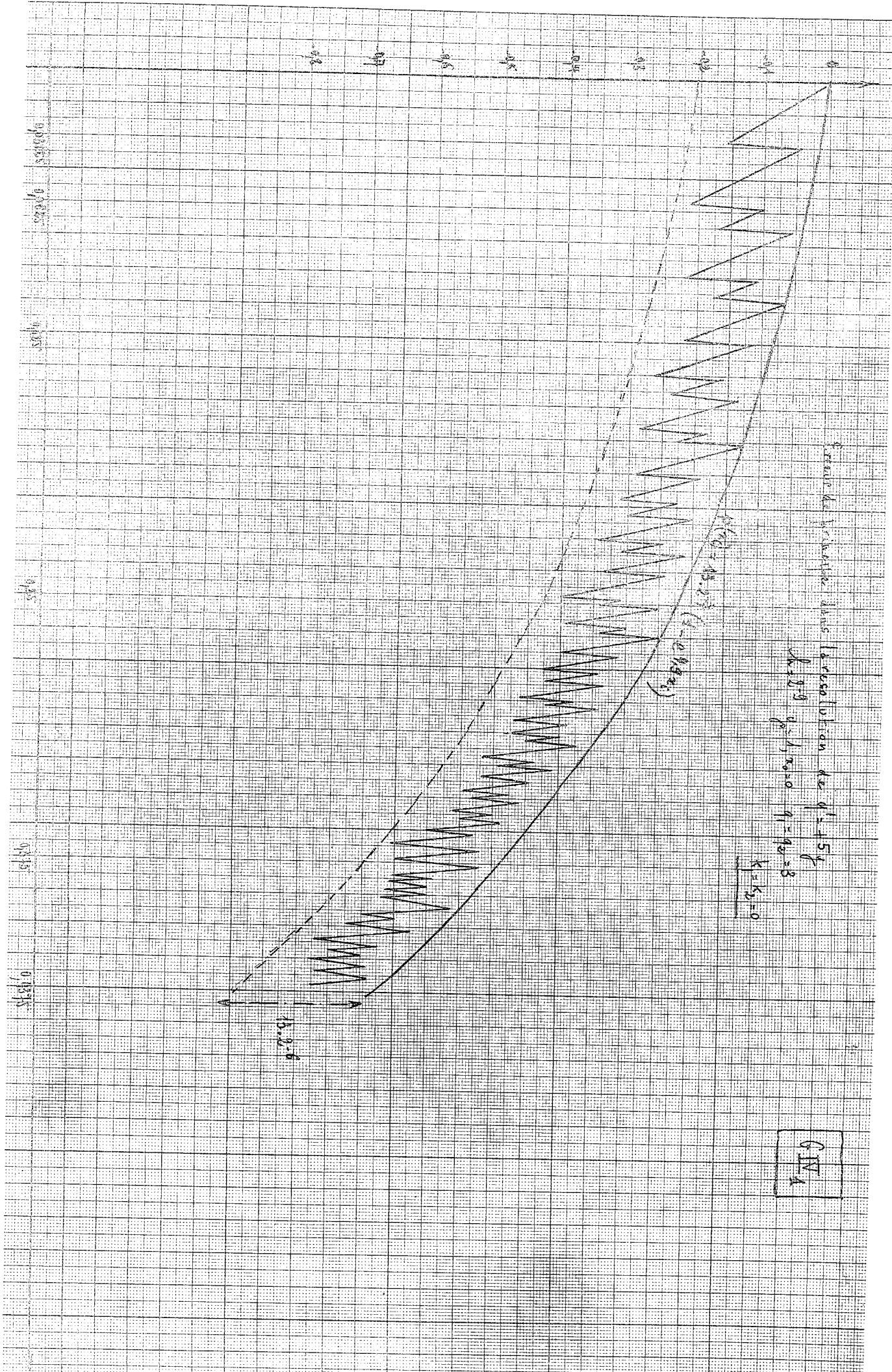
Essai de vibration dans la résolution de $q^i = 1.5f$

$h = 2.9$
 $\rho = 1, m = 0$
 $q = 9.8 = 8$

$k = 5 = 0$

$u(x, t) = 0.2 \sin(\pi x) \cos(2.9 t)$
 $(5 - 0.98 x^2)$

CIV 1

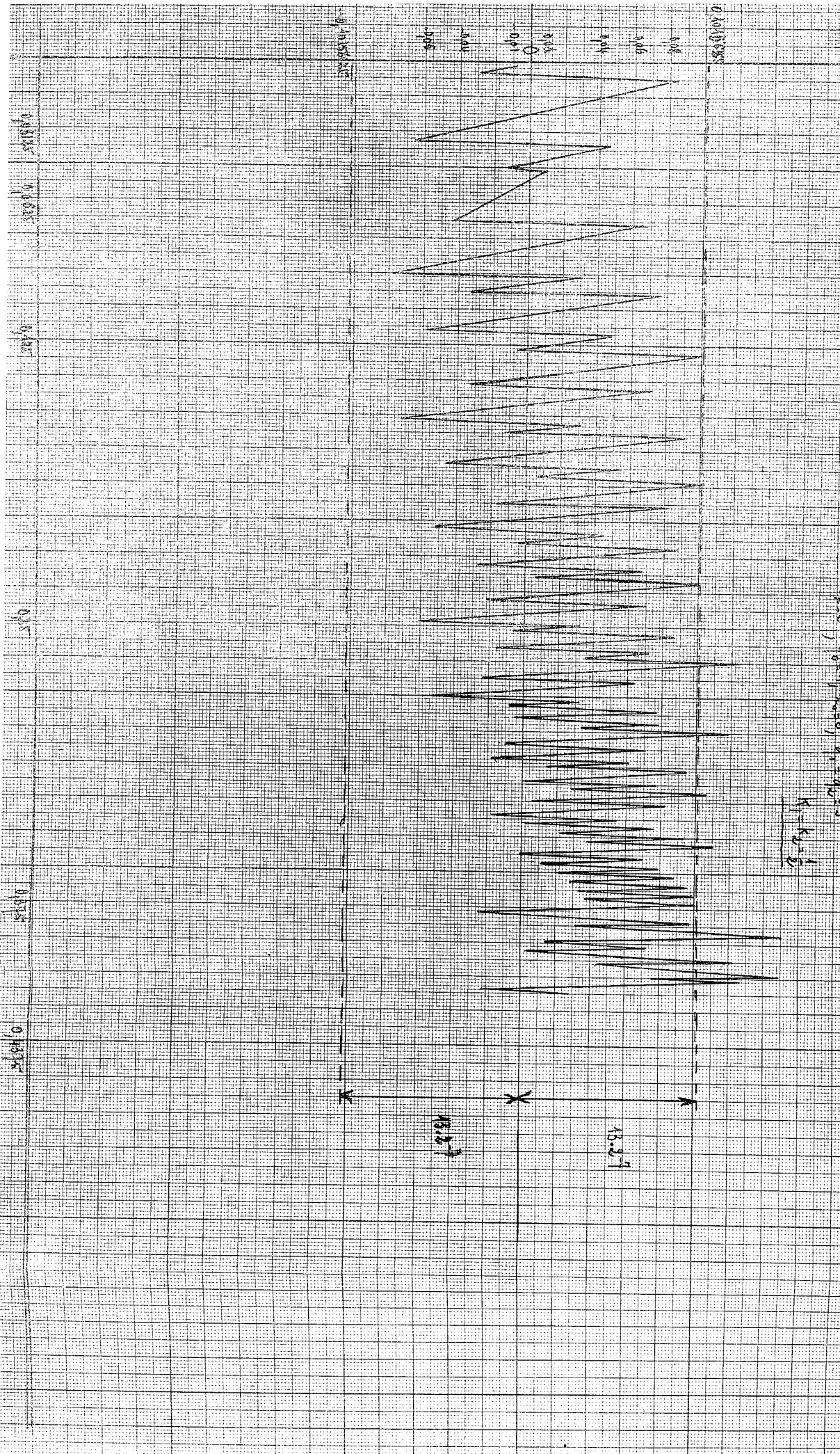


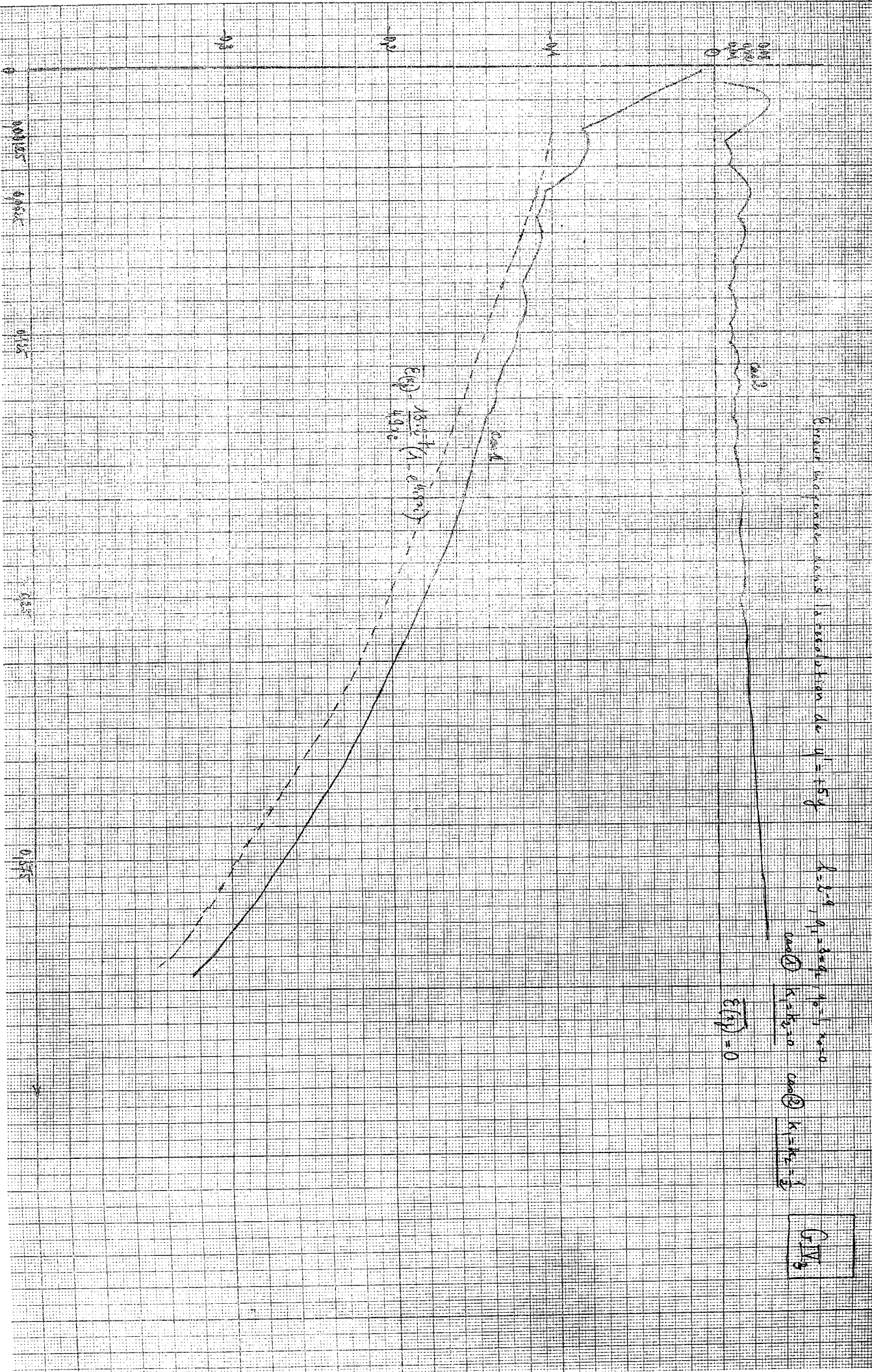
C.IV

Contoh dan penulisan dalam resolusi dan $\mu = 15.5 \mu$

$f = 2.8, 4.0, 1, 2.5, 0, 4, 1, 4, 2.5$

$$k_1 = k_2 = 1$$





$E(\phi) = \frac{18.1}{1 - 0.19\phi^2}$

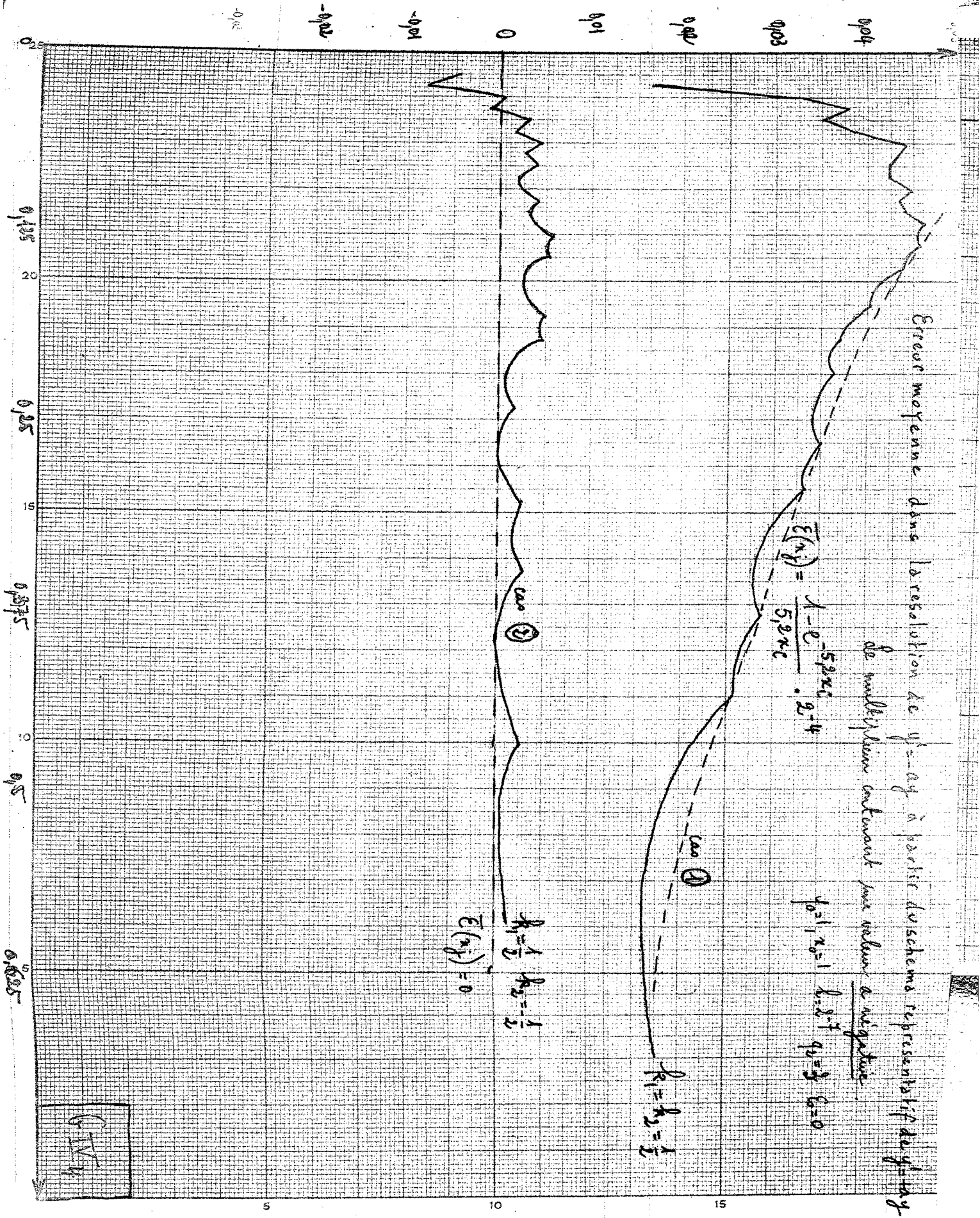
Grundwasserstands-Isolinie

$q = 15 \text{ l/s}$
 $k = 2.4$
 $q_1 = 3 \text{ m}^3/\text{s}$
 $q_0 = 1 \text{ m}^3/\text{s}$

$k = k_2 = 0$
 $k = k_1 = 1$

$0 = (1/3)$

G IV 3



Erreur moyenne dans la resolution de $y = ay$ à partir du schéma représentatif de $y = ay$ de multiplier enlevant pour obtenir a négative.

$$E(ny) = \frac{1 - e^{-5.2\pi x}}{5.2\pi x}$$

$$k_0 = 1, k_1 = 1, k_2 = 2, k_3 = 0$$

$$k_1 = k_2 = \frac{1}{2}$$

$$k_1 = \frac{1}{2}, k_2 = -\frac{1}{2}$$

$$E(ny) = 0$$

GIVE

5

10

15

Variation of structural parameters Q^2

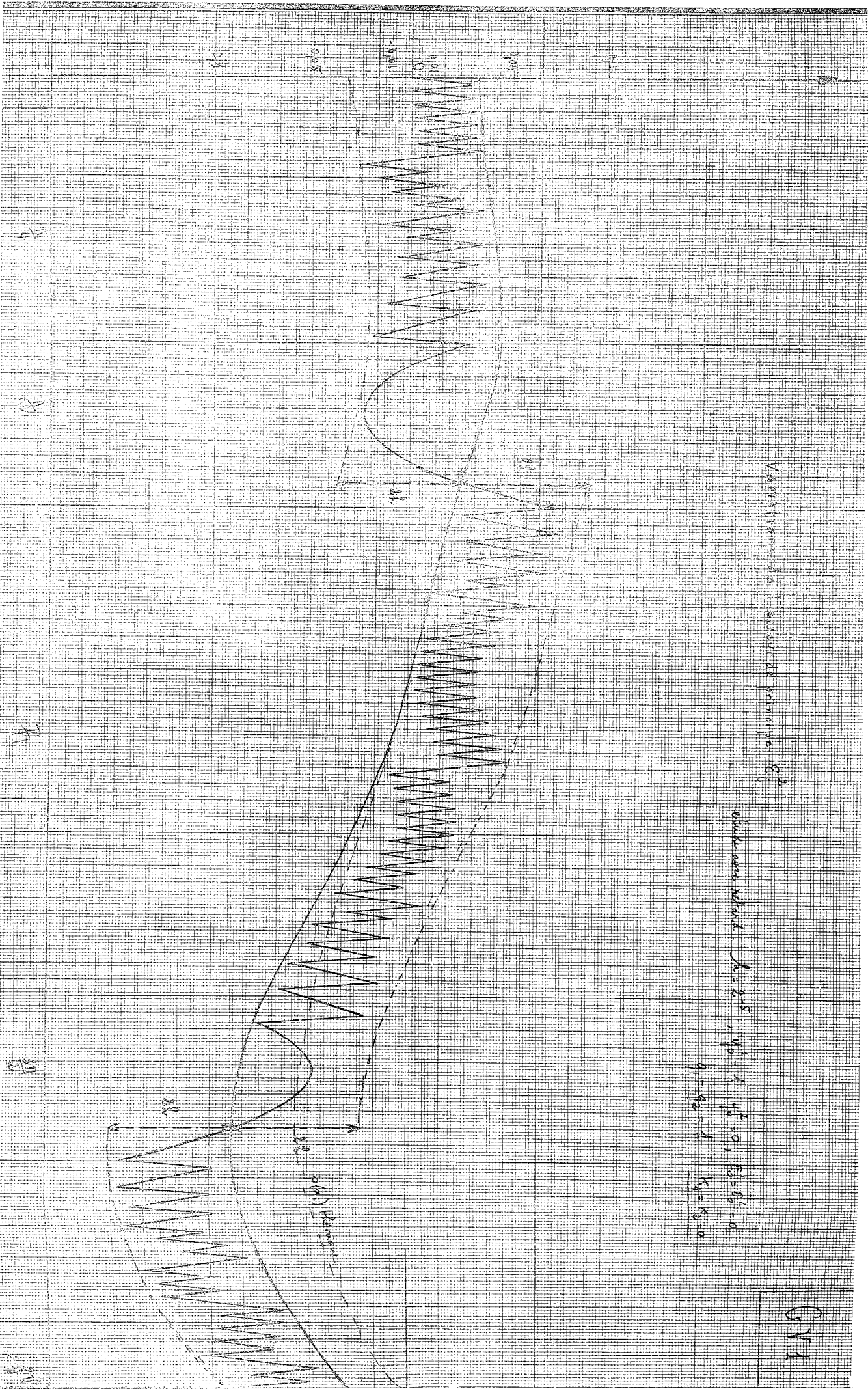
with some related

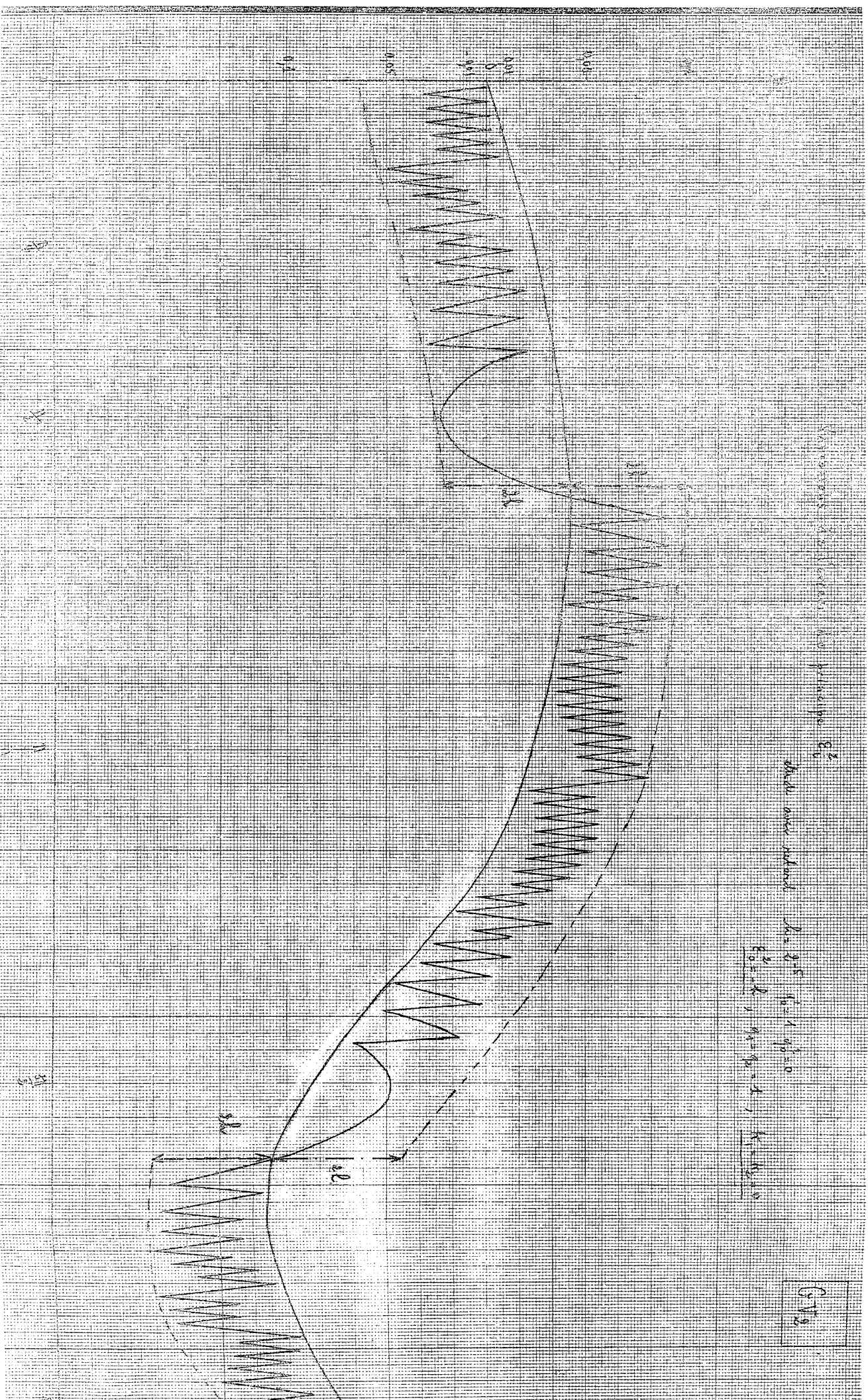
$$A = 2.15, \quad q_1^2 = 1, \quad q_2^2 = 0, \quad E_1^2 = E_2^2 = 0$$

$$q_1 = q_2 = 1$$

$$K_1 = K_2 = 0$$

QV1





Dimensionen in mm
 23
 82

Stahl aus Stahl $\rho_s = 2.5$ $\rho_c = 1.9$

$\rho_s = 2$ $\rho_c = 1.9$

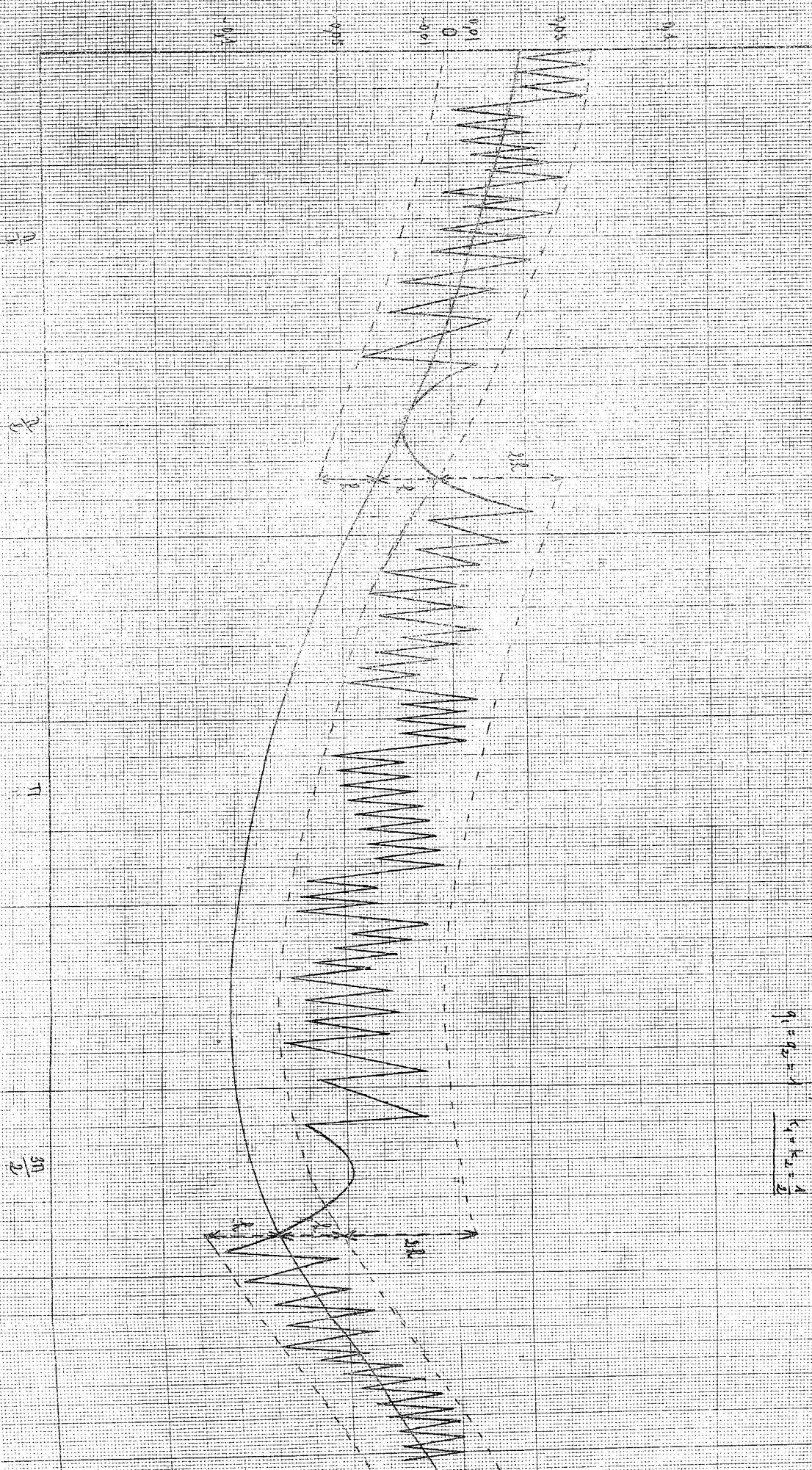
GIV

Variations de l'erreur de principe E₂

Méthode retard

$$\lambda = 2.5 \quad q_0 = 1 \quad q_0^2 = 0$$

$$q_1 = q_2 = 1 \quad k_1 = k_2 = \frac{A}{2}$$

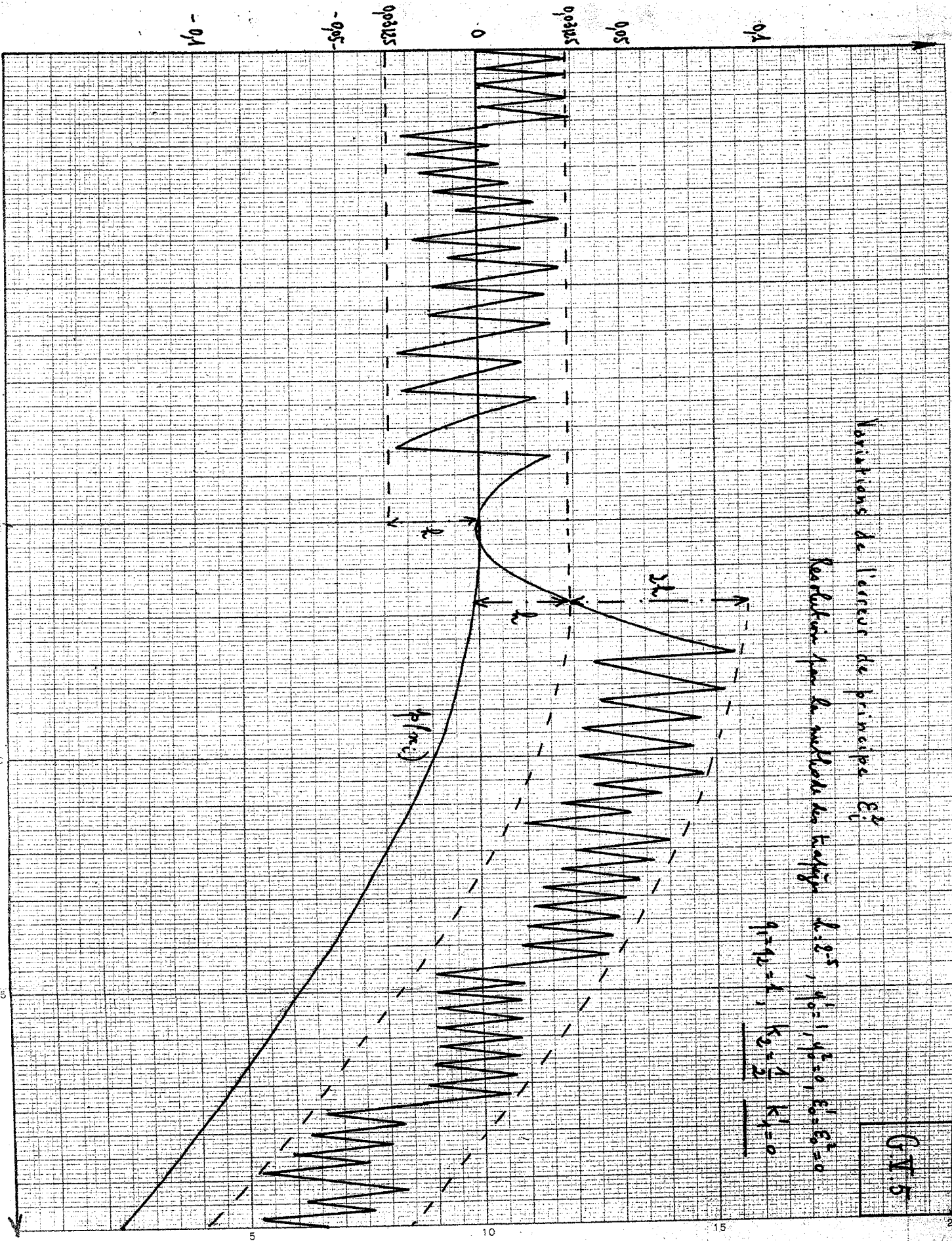


GTS

Variations de l'erreur de principe ϵ_i^p

Recherche pour le meilleur des triangles $\lambda = 2.5$, $\lambda' = 1$, $\lambda'' = 0$, $\epsilon_i^p = \epsilon_i^t = 0$

$\lambda = \lambda' = 1$, $\lambda'' = \frac{1}{2}$, $K_1 = 0$

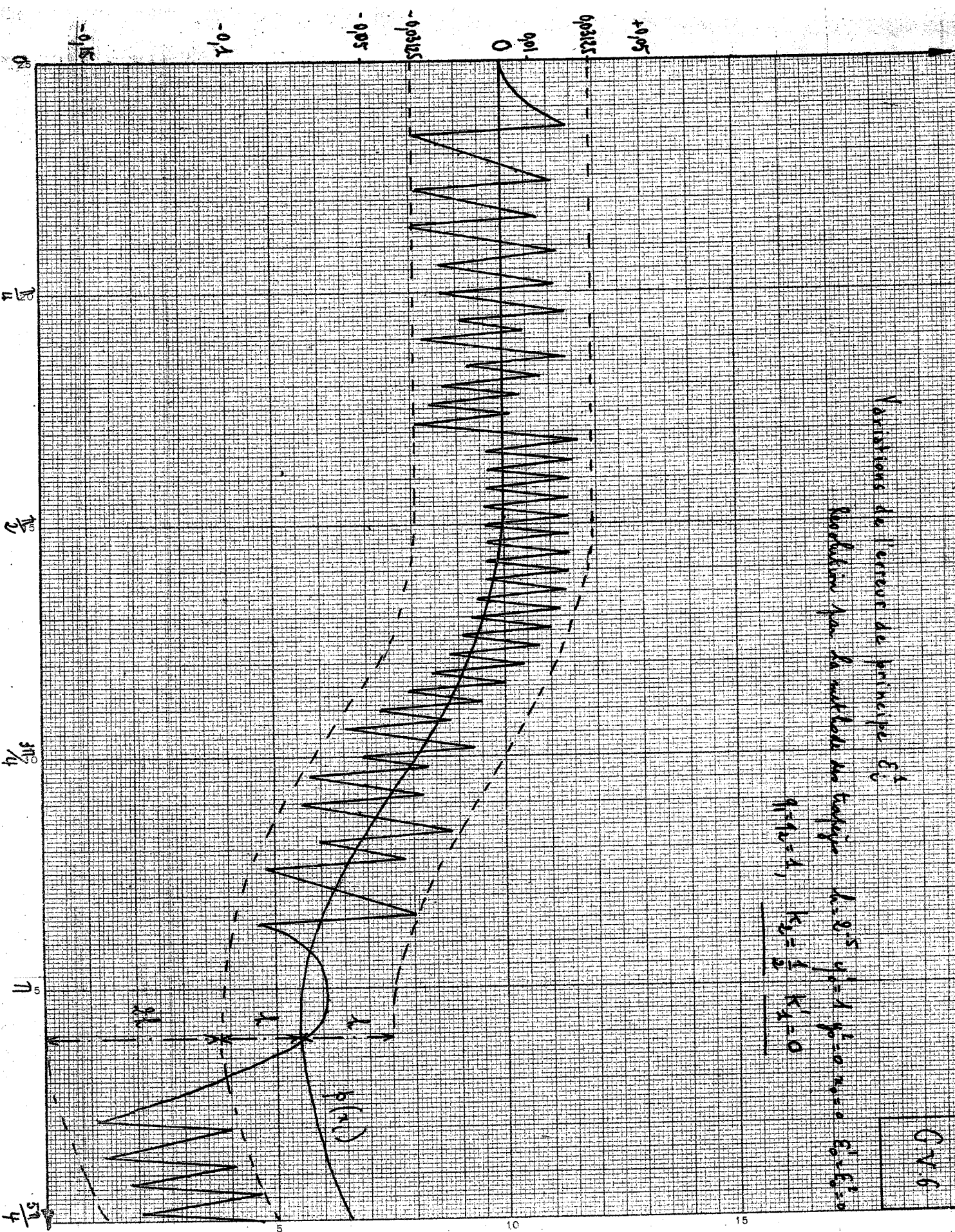


CV6

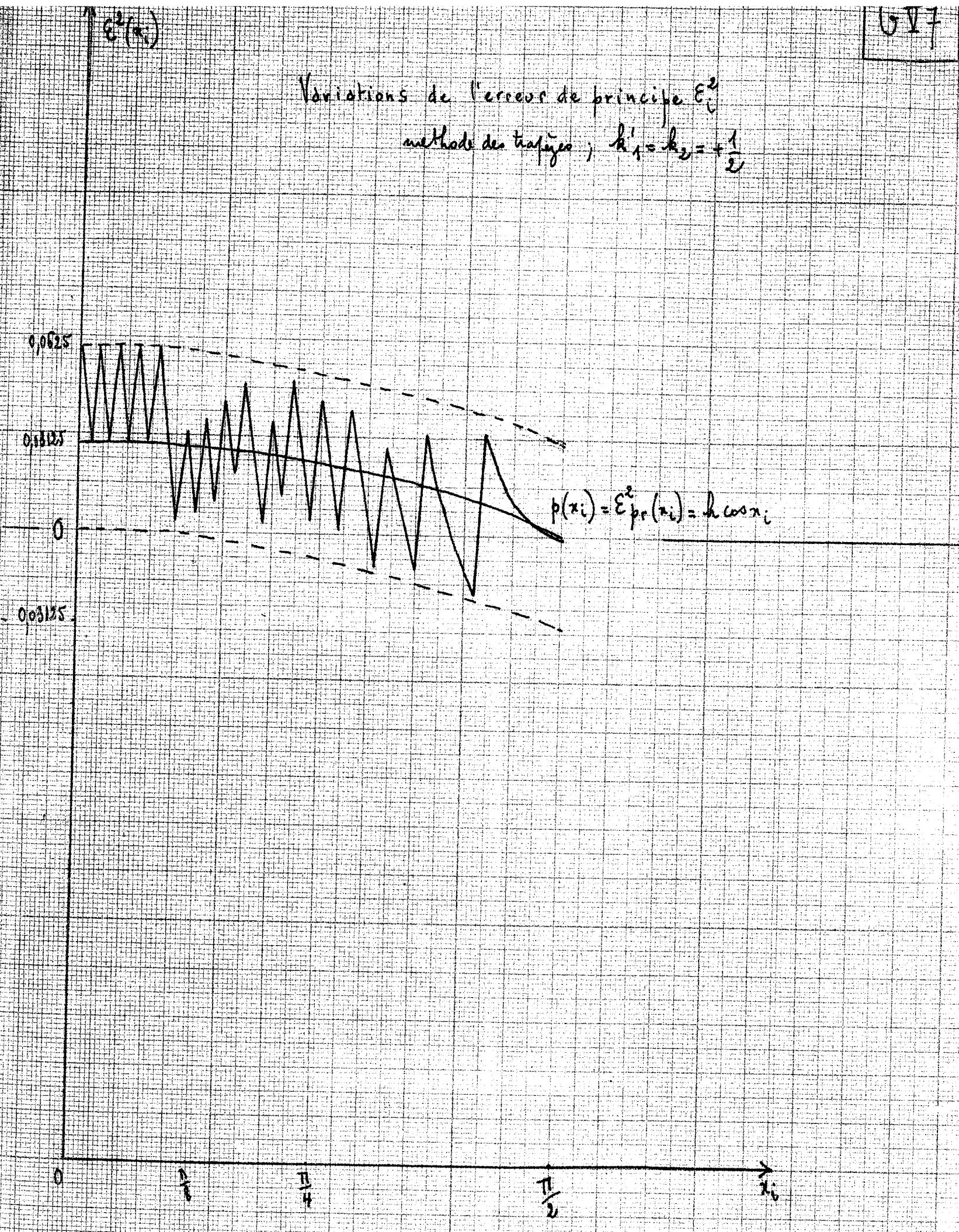
Variations de l'erreur de principe ϵ_p^s

Condition pour la méthode au trapèze $\lambda = 2^5$ $\eta = 1$ $\epsilon = 0$ $\epsilon_1 = 0$ $\epsilon_2 = 0$

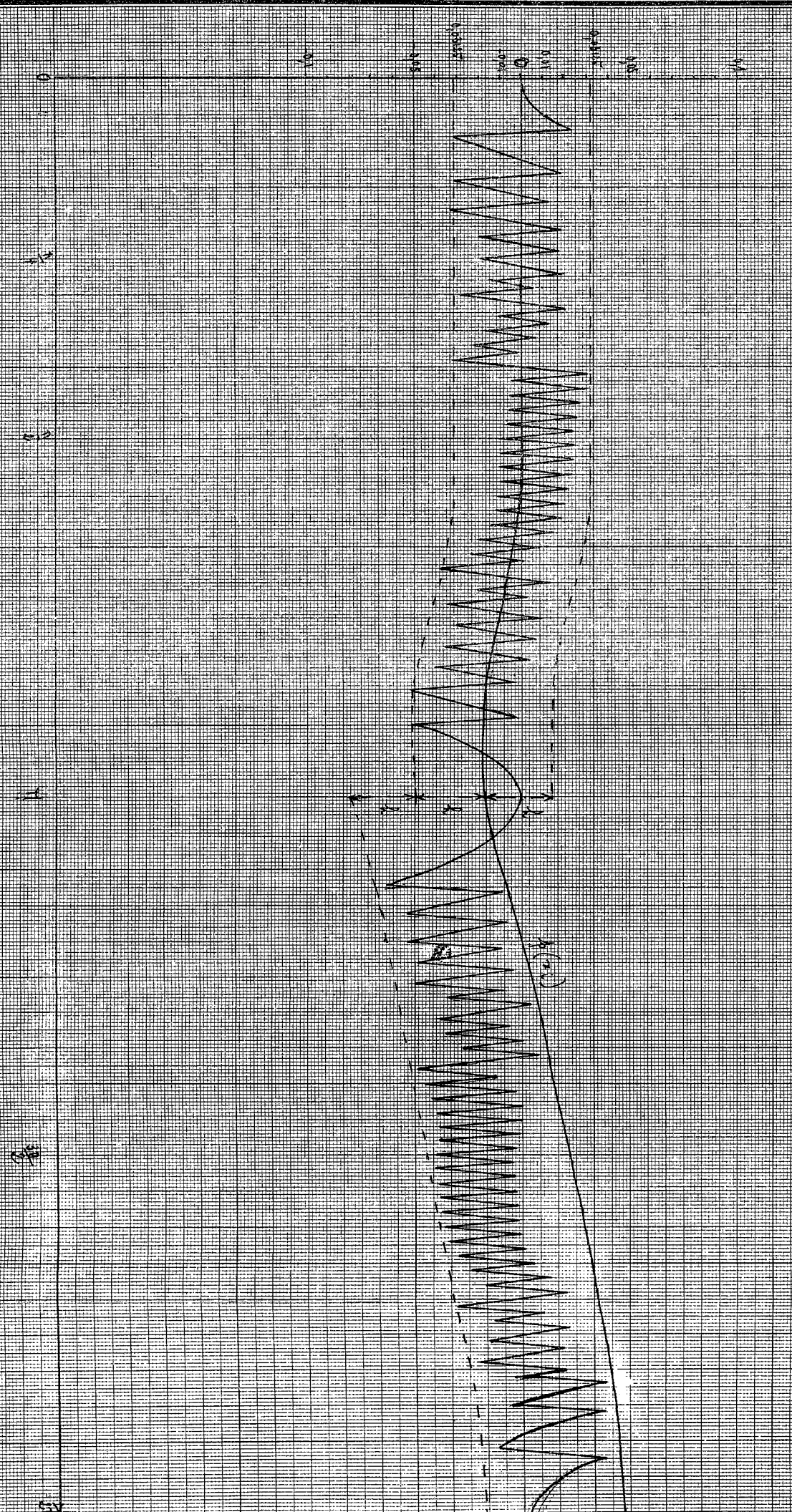
$$q = \eta v = 1, \quad k_2 = \frac{1}{2}, \quad k'_1 = 0$$

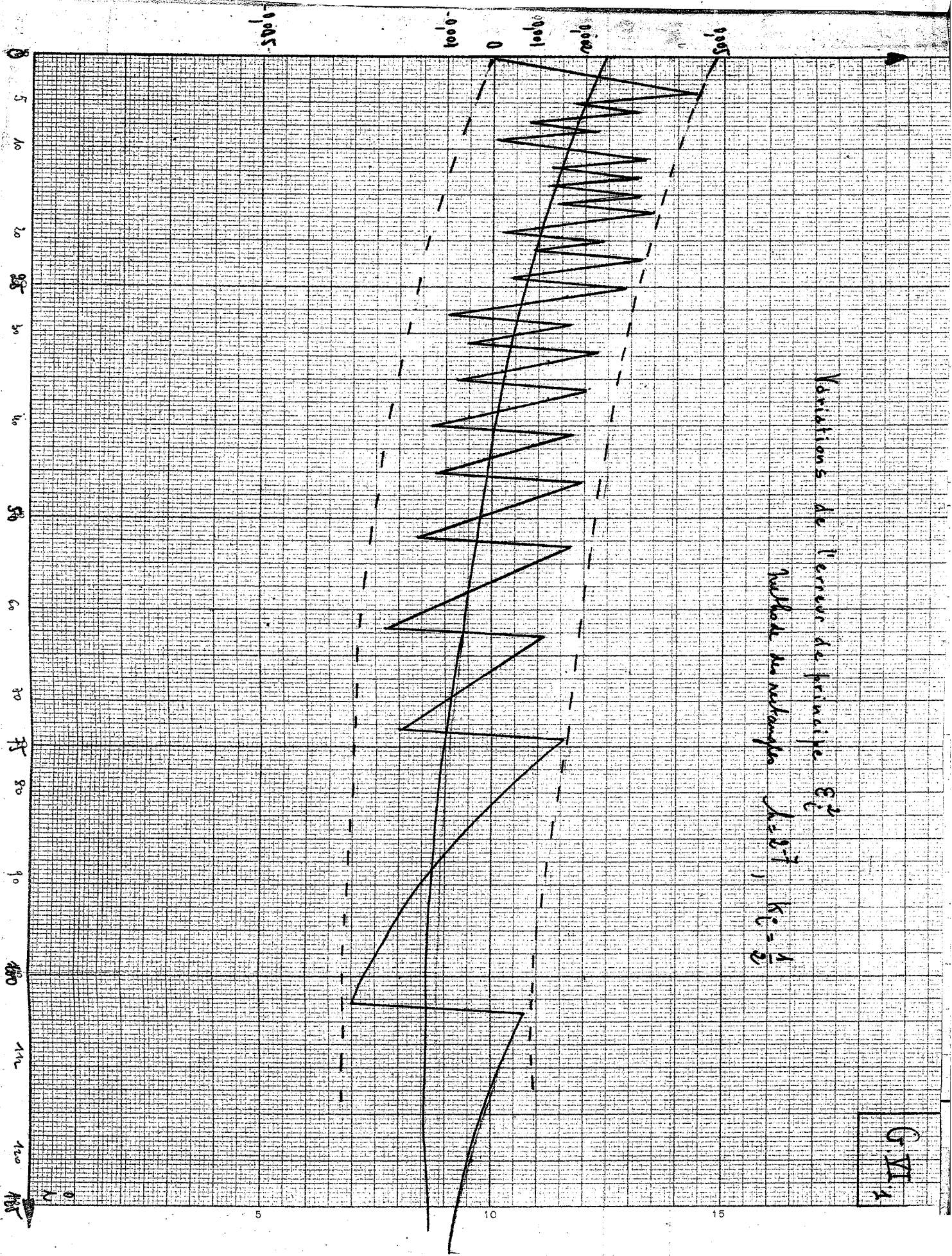


Variations de l'erreur de principe ϵ_i^2
 méthode des trapèzes ; $h_1 = h_2 = \frac{1}{2}$



Modificação do sistema de propulsão
Módulo de energia - K_{10}, K_{11}, K_{12} com a subestrutura





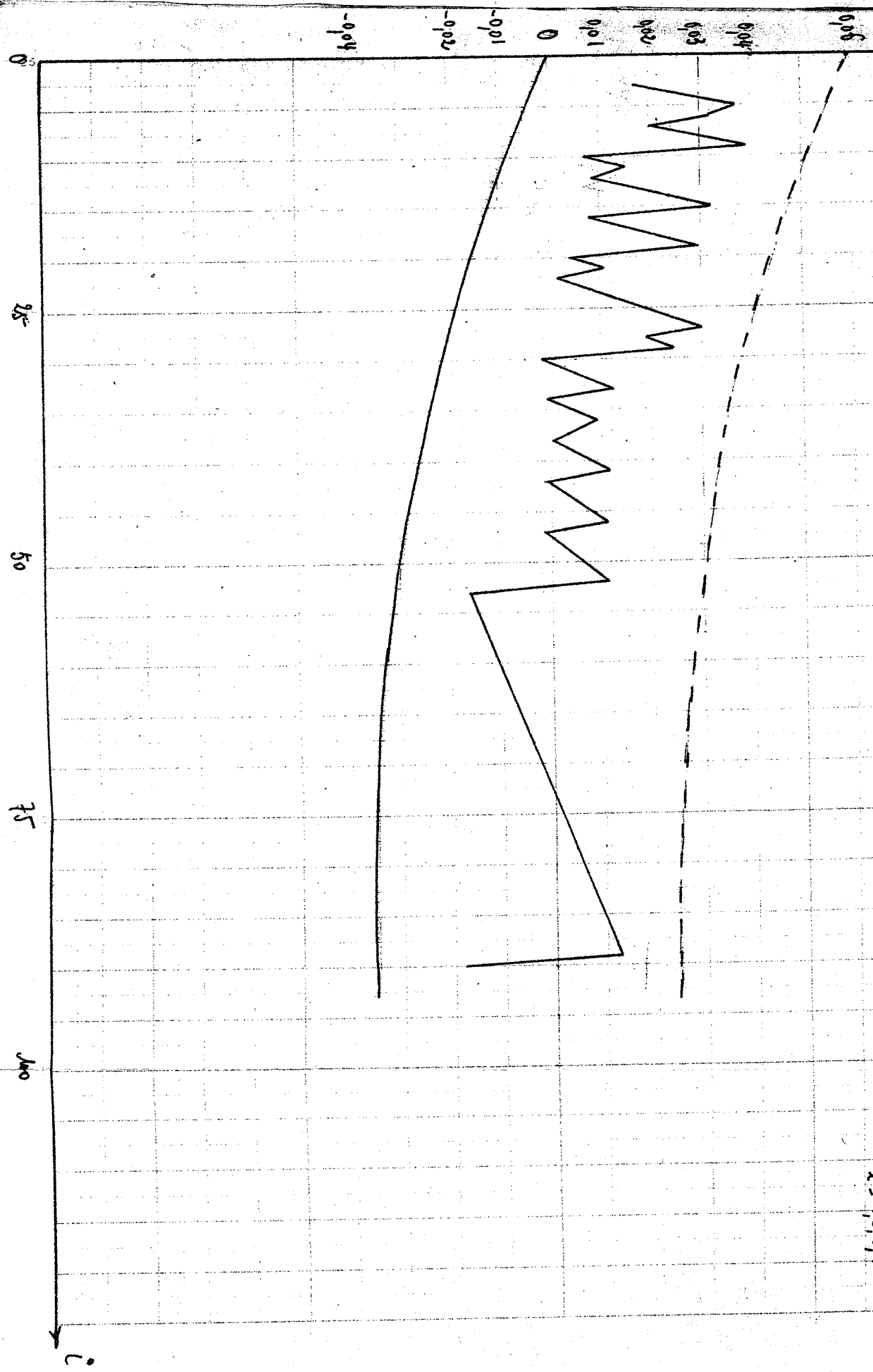
G.VI.1

Variations de l'erreur de principe E_i'

Résolution par la méthode des tangentes

$$K_i = 0$$

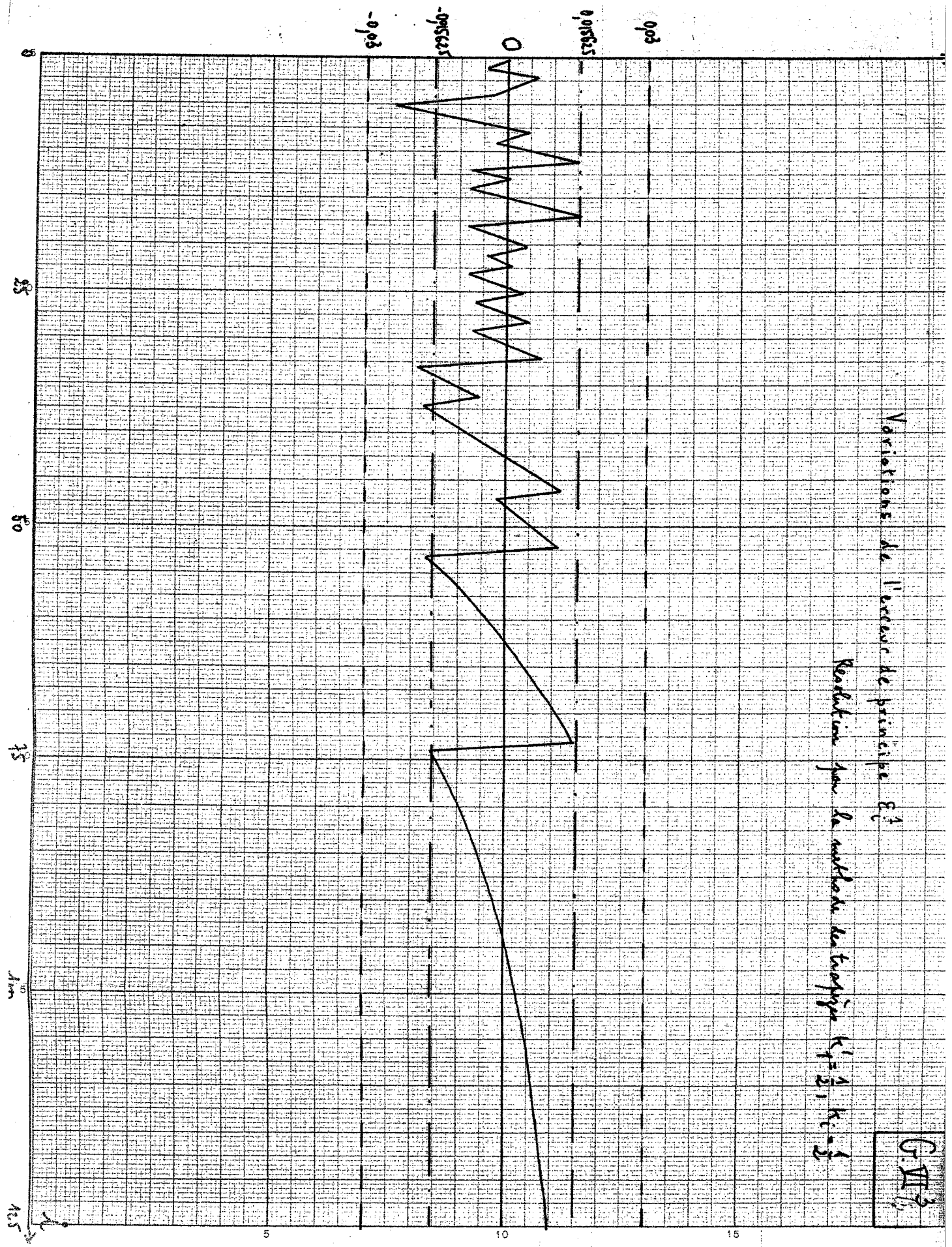
$i = 1, 2, 3, 4$



G.III.3

Variations de l'erreur de principe δ^4

Résultats pour la méthode des trapèzes $K_1=1, K_2=2$

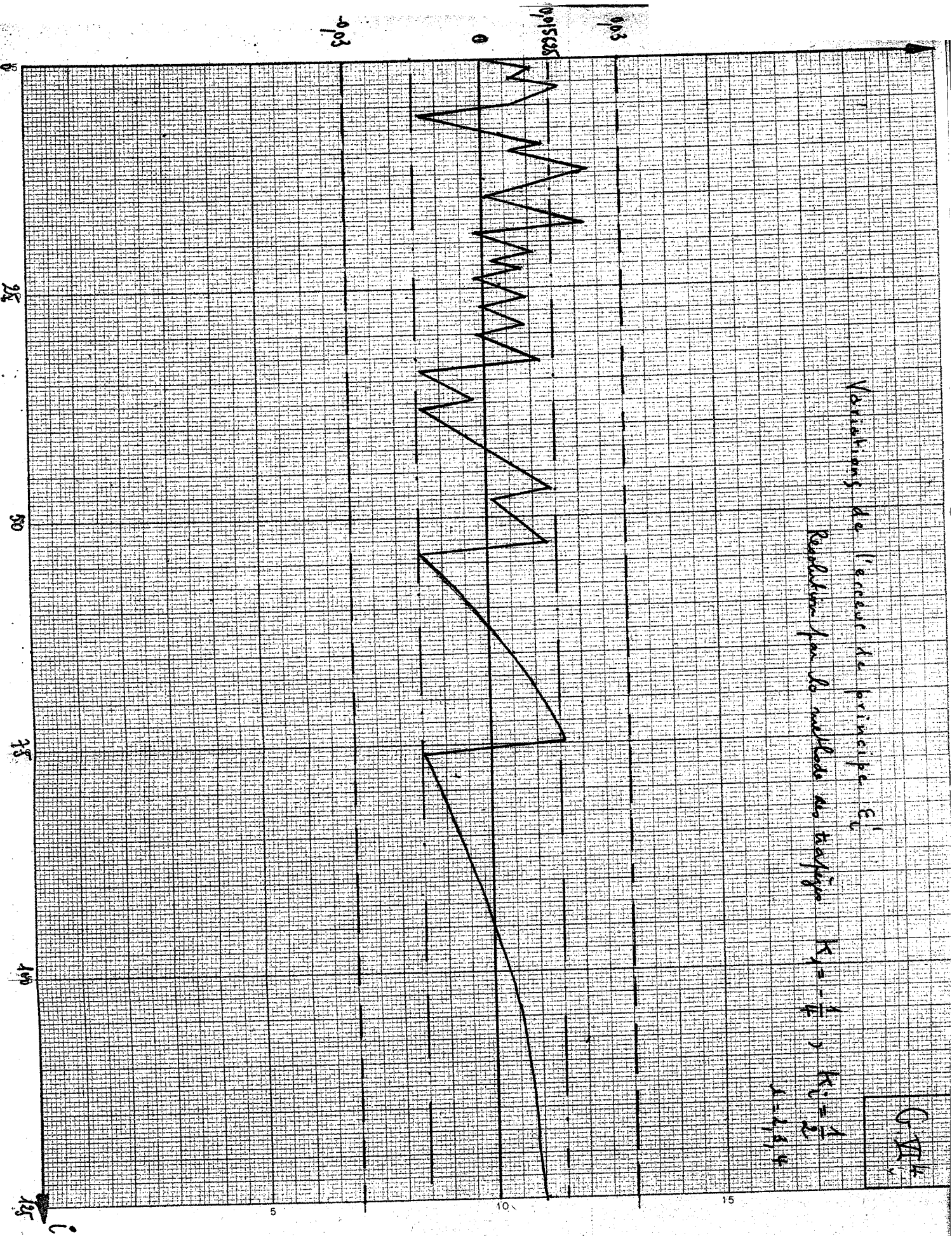


Variations de l'erreur de principe E_1

Règles pour la méthode des tangentes $K_1 = -\frac{1}{4}$, $K_2 = \frac{1}{2}$

$$A = 2,5 \times 10^{-4}$$

G. 211



Variations de l'erreur de principe E_p^2

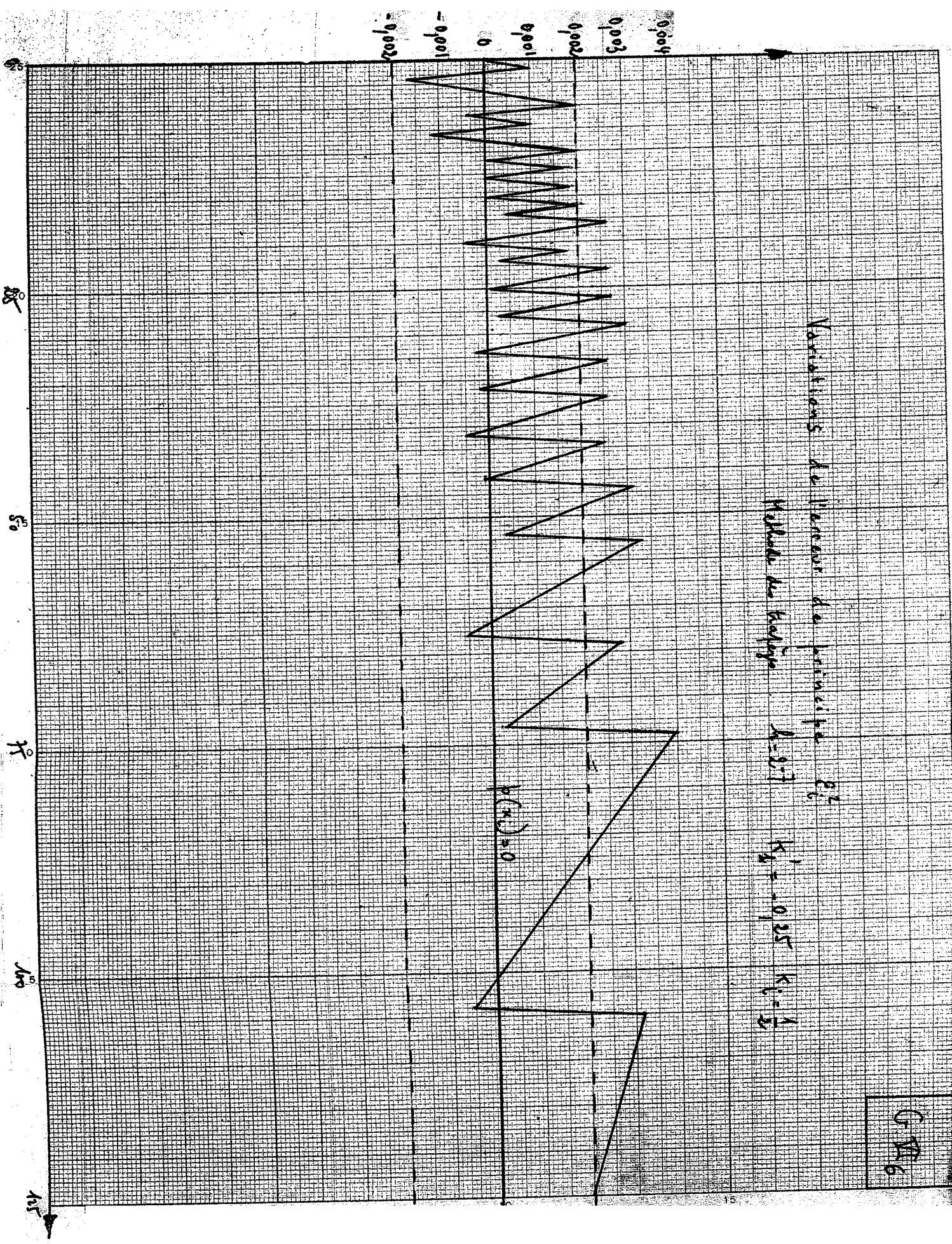
Module du signal $A=2.7$

$K_A=1$

$K_B=-0.25$

$K_C=1$

G 256



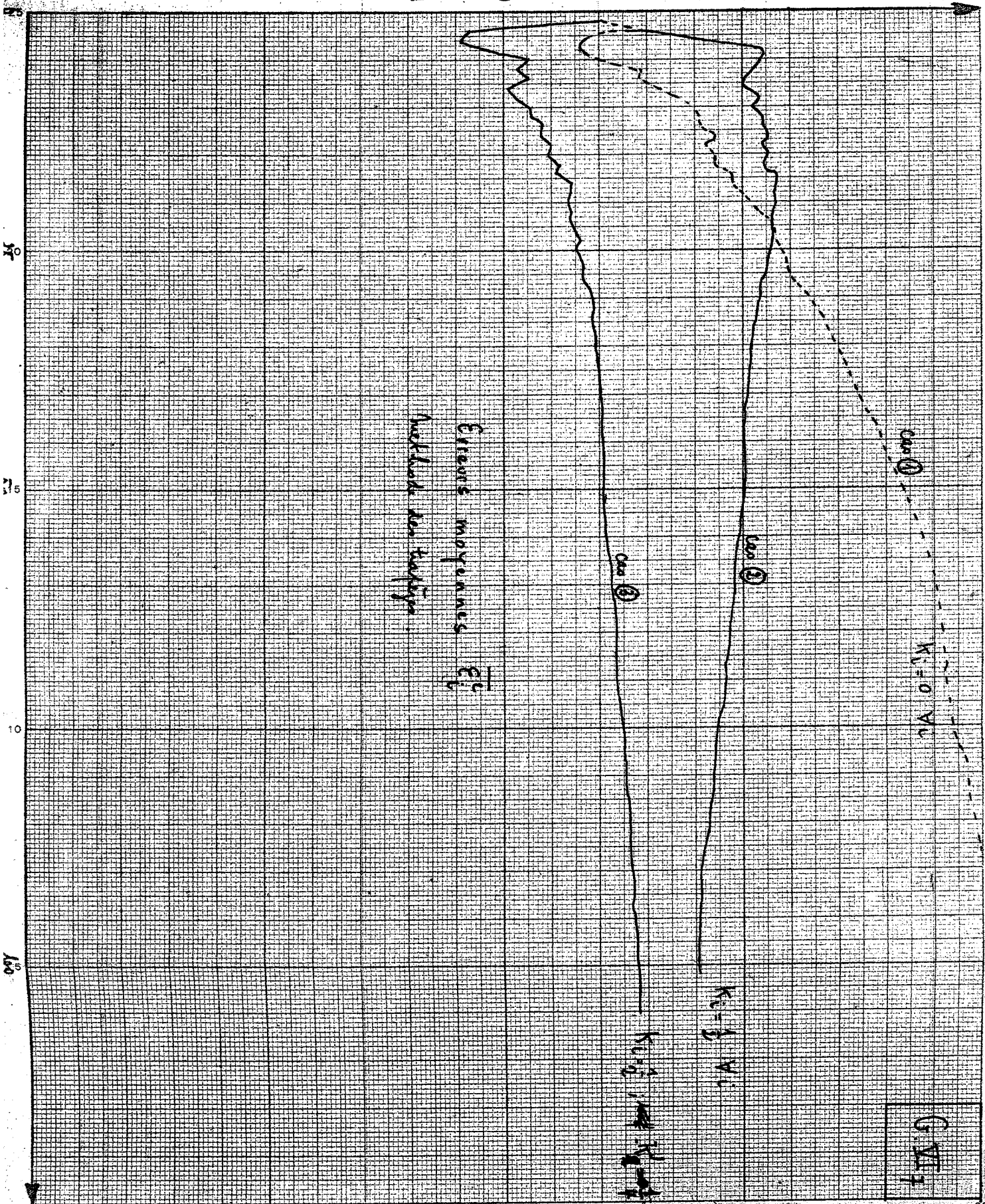


TABLE DES MATIERES

<u>Introduction</u>	<u>pages</u>
A - Rappel du principe de l'analyseur différentiel digital	2
A-1. Principe de l'intégration	2
A-2. Représentation d'un nombre dans un registre ...	3
A-3. Principe de la liaison par incréments	3
A-4. Premières notations	5
B - Exemple du problème $y' = -y$	5
B-1. Schéma représentatif	5
B-2. Relations fondamentales	7
B-3. Erreurs commises	7
C - Généralisation des notations	10
D - But de l'étude	12
CHAPITRE I - Etude de $y' = -y$	
I.A.1 Expression théorique de l'erreur de principe	14
I.A.2 Erreur moyenne dans un intervalle	16
I.A.3 Vérification numérique	16
I.A.4 Erreur de principe et erreur de méthode	17
I.B.1 Réduction de l'erreur de principe	18
I.B.2 Vérification numérique	21
I. C Expression générale de l'erreur de principe remarque : étude de $y' = -y + f(t)$	21
I. D Résolution par la méthode des trapèzes	26

CHAPITRE II	Etude de $y' = + y$	<u>pages</u>
II. A	Etude théorique	28
II. B	Vérification numérique	30
CHAPITRE III	Etude de $y' = - ay$ ($a > 0$)	
III.A.1	Relations fondamentales	33
III.A.2	Expression théorique de l'erreur de principe sur Y_i^1	33
III.A.3	Propagation des arrondis au cours d'un pas de calcul	35
III.A.4	Nouvelle expression de l'erreur de principe. Erreur moyenne	36
III.B	Vérification numérique	37
CHAPITRE IV	Etude de $y' = + ay$ ($a > 0$)	
IV.A.1	Expression théorique de l'erreur de principe..	40
IV.A.2	Vérification numérique	41
IV. B	Multiplication par un nombre négatif ($a < 0$)..	41
PREMIERES CONCLUSIONS		
CHAPITRE V	Résolution de $y'' = - y$	
V.A	Phénomène de retard	45
V.B	Etude de l'erreur de principe dans le cas d'in- tégrations par la méthode des rectangles	50
V.B.1.	Relations fondamentales	50
V.B.2.	Détermination théorique des erreurs de princi- pe	50
V.B.3.	Etude de ϵ_i^2	55
V.B.4.	Vérification numérique	60

	<u>pages</u>
V.C. Intégrations par la méthode des trapèzes	63
V.D. Essai de stabilisation	66
CHAPITRE VI Calcul des racines d'une équation du second degré.	71
VI.A.1 Représentation de la récurrence	71
VI.A.2 Relations fondamentales	73
VI.A.3 Etude théorique des erreurs de principe	74
VI.A.4 Lois de composition des arrondis	81
VI.B Vérification numérique	82
VI.B.1 Méthode des rectangles	83
VI.B.2 Méthode des trapèzes	83
CONCLUSION GENERALE	
ANNEXE.	
Simulation d'un ADD par un ordinateur GAMMA E.T..	88
Caractéristique de l'analyseur simule	89
Description de la simulation	97

VU,

Grenoble, le

Le Président de la Thèse

VU,

Grenoble, le

Le Doyen de la Faculté des Sciences

VU et permis d'imprimer,

Le Recteur de l'Académie de Grenoble