

Algorithmes en lignes, temps-fréquence et apprentissage automatique

Nils Laurent
30 janvier 2025

Notes de cours 2024-2025

Table des matières

1	Introduction	2
2	Conception d'algorithmes efficaces	2
2.1	Complexité algorithmique	2
2.2	Algorithmes en ligne	3
2.3	Compétitivité d'un algorithme en ligne	4
2.4	Références	5
3	Temps-fréquence en ligne	5
3.1	Interpolation linéaire	6
3.2	Choix de la fenêtre d'analyse	7
4	Introduction à l'apprentissage automatique	7
4.1	Apprentissage supervisé	7
4.2	Minimisation du risque empirique	8
4.3	Optimisation des paramètres	9
4.4	Références	9
5	Application : analyse des batteries Li-Ion	10
6	Hors programme	10
6.1	Exhaustivité	10
7	Éléments de base et rappels	11
7.1	Éléments généraux	11
7.2	Probabilités	11
7.3	Intégration	12
7.4	Représentations temps-fréquence	12

1 Introduction

Dans ce cours, nous étudions trois approches qui peuvent être combinées, par exemple pour surveiller un système qui fonctionne en continu. Dans les sections suivantes, une introduction à ces approches est proposée :

- les algorithmes “en ligne” ;
- rééchantillonnage pour l’analyse temps-fréquence ;
- apprentissage automatique.

Chacune d’elle peut s’étudier de manière approfondie, mais cela dépasse le cadre de ce cours, où l’objectif est surtout de donner une intuition autour de la conception d’une méthode. Dans le cadre des travaux pratique, nous nous intéressons à appliquer la méthode vue en cours pour le diagnostic associé au vieillissement d’une batterie de type Li-Ion. Tous les retours (commentaires, suggestions ou erreurs repérées) envoyés à l’adresse suivante sont les bienvenus :

`nils.laurent[AT]univ-st-etienne.fr`

2 Conception d’algorithmes efficaces

Dans le cadre de cette section, nous introduisons des outils d’analyse qui peuvent aider à construire des algorithmes adaptés à lorsqu’un flux continu de données est reçu au cours du temps. Premièrement, nous abordons la notion de complexité d’un algorithme. En suite, une introduction aux algorithmes en ligne est proposée, suivie d’une étude de compétitivité de ce type d’algorithmes.

2.1 Complexité algorithmique

Considérons un algorithme ALG qui prend en entrée une donnée \mathcal{X} de taille N dont nous souhaitons connaître le nombre d’opérations. Par exemple, L’algorithme de recherche du maximum :

Algorithm 1 – Maximum

```
1: Entrée :  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ 
2:  $y = -\infty$ 
3: for  $n = 1, 2, \dots, N$  do
4:   if  $x_n > y$  then
5:      $y = x_n$ 
6:   end if
7: end for
8: return  $y$ 
```

Ce dernier a un nombre de $N + 2$ opération au minimum. Il affecte une valeur à y , fait N comparaisons et affecte le maximum à y . Au maximum, l'algorithme fait $2N + 1$ opération.

Exercice 1 Quelle condition doit on avoir pour atteindre le nombre d'opérations maximal? Même question pour le nombre d'opérations minimal?

La complexité d'un algorithme en fonction de la taille de la variable d'entrée s'étudie généralement au sens de la comparaison asymptotique. On peut par exemple le faire en utilisant la notation de Landau définie ci-dessous.

Définition 1 (Notation de Landau). Soient f, g le coût de deux algorithmes en fonction d'une taille d'entrée N . On note $f = O(g)$ si

$$\exists K \in \mathbb{N}, c > 0, \forall N > K, f(N) \leq cg(N) \quad (1)$$

Souvent, g est un monôme en N d'un certain degré (e.g. $g(N) = N^2$).

Exercice 2 En utilisant la notation de Landau, quelle est la complexité de l'algorithme de recherche du maximum?

2.2 Algorithmes en ligne

Un algorithme "en ligne" est conçu pour fonctionner sur un ensemble de données incomplet, et de prendre en compte les nouvelles données qui lui sont fournies par la suite. Ainsi, il peut se retrouver utile dans des problématiques de diagnostic. Formellement, nous notons $\mathcal{X}_k = \{x_1^k, x_2^k, \dots, x_N^k\} \in \mathbb{R}^N$ la k -ième donnée reçue. Suite à quoi, l'algorithme prend en compte les données reçues avant l'itération k , dans une variable de synthèse Y_{k-1} . Pour comprendre l'intérêt examinons deux extrêmes :

- Si Y_{k-1} contient toutes les données précédentes, alors on a le plus d'information possible mais l'algorithme sera probablement gourmand en espace mémoire et en nombre d'opérations.
- Si Y_{k-1} est vide, alors l'information précédente est ignorée.

On souhaite donc définir Y_{k-1} de sorte à contenir le plus d'informations sans être trop volumineux.

L'algorithme ci-dessous met en évidence les étapes d'un tel algorithme.

Algorithm 2 – Algorithme en ligne

- 1: Réception de \mathcal{X}_k
 - 2: $Y_0 = \emptyset$
 - 3: Calculer Y_k en fonction de \mathcal{X}_k et Y_{k-1}
 - 4: Mettre à jour Y_{k-1} pour l'itération suivante
 - 5: Calculer le résultat z en fonction de Y_k et \mathcal{X}_k
-

Exercice 3 Soient les variables aléatoires X_1, X_2, \dots, X_k qui suivent toutes la loi de Bernoulli $\mathcal{B}(\theta)$ où θ est le paramètre de la loi et qui prennent des valeurs dans $\{0, 1\}$.

- 1) Rappeler quelle est l'espérance et la variance pour une loi de Bernoulli $\mathcal{B}(\theta)$.
- 2) Quelle statistique permet d'estimer la loi de X_1, X_2, \dots, X_k ?
- 3) Écrire un algorithme en ligne qui prend une variable à la fois et qui estime la loi suivie par X_1, X_2, \dots, X_k .

2.3 Compétitivité d'un algorithme en ligne

Pour évaluer la performance d'un algorithme $\text{ALG} : \mathbb{R}^N \rightarrow Z$, nous introduisons une fonction de coût $J : Z \rightarrow \mathbb{R} \cup \{\infty\}$ continue. Il est en suite recherché l'élément $z \in Z$ qui minimise le cout donné par J , noté z^* , i.e.

$$z^* \in \underset{z \in F}{\text{Arg min}} J(z). \quad (2)$$

Définition 2 (Algorithme c -compétitif, cas déterministe). Soit ALG un algorithme en ligne, et OPT l'algorithme qui donne la solution optimale. Soit $c > 0$, ALG est dit c -compétitif si $\exists b \in \mathbb{R}$ tel que $\forall x \in \mathbb{R}^N$

$$J(\text{ALG}(x)) \leq cJ(\text{OPT}(x)) + b = cJ(z^*) + b.$$

Définition 3 (Ratio compétitif). Le ratio compétitif noté par c_{ALG} est défini de la manière suivante

$$c_{\text{ALG}} := \inf_{c > 0} c \quad \text{tel que ALG est } c\text{-compétitif.} \quad (3)$$

Exercice 4 Supposons que nous devons suivre le bon fonctionnement d'une machine et gérer son coût de maintenance. Au total, nous avons N observation x_1, \dots, x_N . Une observation x_k vaut f si l'état est

fonctionnel, d si il est défaillant. Le coût de maintenance associé à l'état d est C_d . Il est possible de sous-traiter à n'importe quel indice k pour un coût $C_s = MC_d$, au quel cas, la maintenance associée aux prochains états (i.e. x_{k+1}, \dots, x_N) est à la charge du sous-traitant.

- a) Quel est le coût de l'algorithme optimal ?
- b) Considérons l'algorithme en ligne suivant :

```

Recevoir  $x_k$ 
 $Y_0 = (0, 0)$ 
if  $Y_{k-1,0} \geq M/2$  then
     $Y_k = (Y_{k-1,0}, 1)$ 
else if  $x_k == d$  then
     $Y_k = (Y_{k-1,0} + 1, 0)$ 
end if
return  $Y_{k,0}C_d + Y_{k,1}C_s$ 

```

Trouver une constante c et montrer que l'algorithme est c -compétitif.

2.4 Références

François Schwarzenruber, "Algorithmes en ligne", https://people.irisa.fr/Francois.Schwarzenruber/algo2/08onlinealgo_student.pdf

3 Temps-fréquence en ligne

Pour faire une analyse temps-fréquence, nous allons supposer que la k -ième donnée reçue contient le signal x et l'information des instants de temps associées aux mesures.

Formellement, les données prennent la forme suivante

$$\mathcal{X}_k = \{(x_1, t_1), (x_2, t_2), \dots, (x_L, t_L)\}, \quad (4)$$

où pour $n \in \{1, \dots, N\}$, $x_n \in \mathbb{R}$ représente la valeur du signal x aux temps $t_n \in \mathbb{R}$.

Maintenant, nous souhaitons obtenir la représentation du signal dans l'espace d'analyse de la transformée de Fourier à court terme (TFCT). Pour cela, il nous faut introduire les paramètres nécessaires, c'est-à-dire

- N_{fft} : nombre de coefficients souhaités sur l'axe des fréquences
- $g \in \mathbb{R}^M$: la fenêtre d'analyse

A condition que les mesures soient faites sur des pas de temps uniformes, la représentation temps-fréquence obtenus est simplement

$V_x^g \in \mathbb{R}^{L \times N_{fft}}$. Maintenant, si les pas ne sont pas uniformes, on peut procéder à du rééchantillonnage, mais il faut définir ce que l'on entend par là.

Rééchantillonner, c'est estimer une mesure qui n'a pas été acquise : nous ne connaissons rien en dehors des instants t_1, \dots, t_L . Une manière de compléter l'acquisition est de faire de l'interpolation.

3.1 Interpolation linéaire

Nous allons considérer un des cas les plus simples et utiliser un polynôme d'interpolation linéaire (donc de degré ≤ 1). Un tel polynôme s'écrit

$$h(t) = a_1 + a_2 t \quad (5)$$

où $a_1, a_2 \in \mathbb{R}$. Nous pouvons par la suite interpoler x_1, x_2 en t_1, t_2 respectivement via la résolution du système

$$\begin{cases} a_1 + a_2 t_1 = x_1 \\ a_1 + a_2 t_2 = x_2 \end{cases} \quad (6)$$

Remarque 4. Un polynôme d'interpolation peut avoir un degré plus grand que 1, même si cela n'est pas traité dans ce cours. Il est important, lorsque l'on considère un problème d'interpolation, d'étudier le phénomène de Runge et d'être conscient de la potentielle erreur que l'on peut avoir si on interpole naïvement avec un degré élevé. Cependant, il est possible de gagner en précision en augmentant le degré, par exemple en utilisant les racines de Chebyshev.

Avec ce polynôme, nous savons interpoler deux valeurs, mais il nous faut l'étendre pour l'appliquer à tout le signal. Une méthode possible est d'utiliser une spline, ce que nous introduisons maintenant.

Définition 5. Une spline de degré $d \geq 1$ interpolant x sur $[t_1, t_L]$ aux points x_1, x_2, \dots, x_L est la fonction h de classe $C^{d-1}(\mathbb{R})$ vérifiant $\forall n \in \{1, \dots, L\}, h(t_n) = x_n$ et pour $n < L - 1$ sa restriction sur $[t_n, t_{n+1}]$ est un polynôme de degré d .

Dans notre cas, $h \in C^0(\mathbb{R})$ est simplement la fonction affine par morceaux qui passe par x_1, x_2, \dots, x_L .

Exercice 5 Considérons que suite à une mesure, on a accès à $\mathcal{X}_k = \{(x_1, t_1), (x_2, t_2), \dots, (x_L, t_L)\}$.

a) Proposer un rééchantillonnage de x à partir d'une spline h de sorte que les échantillons soient uniformément espacés en temps.

b) Rappelons que la fonction $\text{sinc}(\pi t) = \frac{\sin(\pi t)}{\pi t}$ est un filtre passe bas exprimé dans le domaine temporel. Quel domaine du spectre est filtré par cette fonction ?

c) Supposons que le signal en entrée est trop volumineux. Comment peut-on le rééchantillonner pour obtenir un signal deux fois plus léger ?

d) Lorsque l'on sous-échantillonne pour rendre le signal plus léger, est-ce que l'on perd de l'information ?

3.2 Choix de la fenêtre d'analyse

Nous avons supposé que chaque donnée reçue, notée \mathcal{X}_k , contient l'information de L échantillons du signal x . Nous allons aussi supposer que le signal est échantillonné uniformément en temps, car il est possible de se ramener à ce cas par interpolation spline (même si on a alors une approximation de la mesure).

Pour analyser le signal, il est important de choisir une fenêtre g adaptée. Dans le cadre continu, considérons la fenêtre gaussienne suivante

$$g_\sigma(t) = e^{-\pi \frac{t^2}{\sigma^2}}. \quad (7)$$

Exercice 6 Étude du choix de la fenêtre.

a) A partir de σ , comment obtient-on une meilleure localisation temporelle/fréquentielle ?

b) Lorsque $\sigma \gg 0$, à quel problème sommes-nous confronté lorsque nous souhaitons analyser \mathcal{X}_k ? (Indice : étudier le cas où la fenêtre est centrée en temps).

4 Introduction à l'apprentissage automatique

L'apprentissage automatique est une approche qui s'inscrit dans la science des données et l'intelligence artificielle. Il s'appuie sur :

- les statistiques ;
- l'optimisation ;
- l'algorithmique ;
- le traitement du signal.

Cet apprentissage repose sur les données (que l'on suppose avoir en quantité suffisante) et l'algorithme d'apprentissage. La science des données va s'intéresser à la préparation des données et de sa représentation. L'algorithme d'apprentissage a pour objectif de choisir les bons paramètres du modèle afin de traiter de nouvelles données.

4.1 Apprentissage supervisé

L'apprentissage supervisé est un cadre dans lequel à partir du jeu de données \mathcal{X} , les paramètres θ du modèle ϕ_θ sont calibrés en utilisant

les solutions \mathcal{Y} associées à \mathcal{X} . Suite à l'entraînement, ϕ_θ doit apprendre à partir de $x \in \mathcal{X}$ à faire une prédiction. Formellement, cela s'exprime par la définition suivante.

Définition 6 (Apprentissage supervisé). Soient $\{x_n\}_{n=1,\dots,N} \subset \mathcal{X}$ les données d'entraînement, $\{y_n\}_{n=1,\dots,N} \subset \mathcal{Y}$ l'ensemble des étiquettes des données. On suppose qu'il existe une fonction $\phi : \mathcal{X} \rightarrow \mathcal{Y}$ tel que $y_i = \phi(x_i) + \varepsilon_i$, où ε_i est un bruit aléatoire. L'objectif de l'apprentissage est de trouver les paramètres θ tels que la fonction $\phi_\theta : \mathcal{X} \rightarrow \mathcal{Y}$ vérifie $\phi_\theta(x) \approx \phi(x)$ pour tout couple $(x, \phi_\theta(x)) \in \mathcal{X} \times \mathcal{Y}$.

Dans la définition précédentes remarquons que :

- ε_i est introduite pour prendre en compte les erreurs de mesures qui peuvent être présentes dans les étiquettes ;
- la relation entre ϕ_θ et ϕ est vérifié sur l'espace \mathcal{X} car nous souhaitons que ϕ_θ généralise à partir des données d'entraînement.

Définition 7 (Classification binaire). Un problème d'apprentissage est dit de classification binaire si les étiquettes $y \in \mathcal{Y}$ prennent des valeurs dans $\{0, 1\}$.

Exercice 7 Décrire un problème de classification binaire en traitement d'images en utilisant le formalisme ci-dessus.

4.2 Minimisation du risque empirique

Un problème de minimisation fait intervenir une fonction de coût. Dans le contexte de l'apprentissage supervisé, nous considérons la définition suivante.

Définition 8 (Fonction de coût). Une fonction de coût $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ est une fonction utilisée pour quantifier la qualité d'une prédiction. Plus la valeur du coût $\mathcal{L}(y, \phi_\theta(x))$ est faible, plus $\phi_\theta(x)$ sera considérée de bonne qualité.

Puis, on peut définir le risque empirique de la façon suivante.

Définition 9 (Risque empirique). Dans un cadre supervisé, considérons les N observations étiquetées $\{(x_n, y_n)\}_{n=1,\dots,N}$, on appelle le risque empirique de ϕ_θ la quantité suivante

$$R_N(\phi_\theta) = \frac{1}{N} \sum_{n=1}^N \mathcal{L}(y_n, \phi_\theta(x_n)) \quad (8)$$

Exercice 8 Risque empirique et classification binaire

a) Supposons que suite à l'entraînement, $P[\phi_\theta(x_n) = y_n] = \alpha$, où $\frac{1}{2} < \alpha < 1$. En moyenne, quelle va être le risque empirique ?

b) Si $P[\phi_\theta(x_n) = y_n] \leq \frac{1}{2}$, la méthode a-t-elle un intérêt ?

4.3 Optimisation des paramètres

La branche de l'optimisation est essentielle à la conception d'algorithmes d'apprentissage. Par exemple, l'optimisation peut être appliquée au choix du paramètres θ pour la fonction ϕ_θ introduite précédemment. Dans cette partie, nous allons étudier l'algorithme de la descente de gradient, qui est l'une des méthodes d'optimisation les plus simples. De plus, nous allons mener cette étude sur une partie des fonctions $J \in C^1(\mathbb{R}, \mathbb{R})$, qui vérifient certaines propriétés permettant de bien poser le problème d'optimisation.

Tout d'abord, nous supposons que J est convexe.

Définition 10 (Convexité d'une fonction). La fonction $J : \mathbb{R} \rightarrow \mathbb{R}$ est convexe, si pour $x, y \in \mathbb{R}^N$, $t \in]0, 1[$

$$J((1-t)x + ty) \leq (1-t)J(x) + tJ(y). \quad (9)$$

Exercice 9 Trouver une fonction convexe qui n'a pas de minimum.

Supposons aussi que J est coercive.

Définition 11 (Coercivité). La fonction $J : \mathbb{R} \rightarrow \mathbb{R}$ est coercive si

$$\lim_{x \rightarrow +\infty} J(x) = \lim_{x \rightarrow -\infty} J(x) = +\infty \quad (10)$$

Si la fonction est coercive, alors il existe x^* tel que $J(x^*)$ est l'unique minimum.

Nous allons également supposer que J' est β -Lipschitz

Définition 12 (Lipschitz continuité). Une fonction $G : \mathbb{R} \rightarrow \mathbb{R}$ est β -Lipschitz continue sur \mathbb{R} si pour $x, y \in \mathbb{R}$, elle vérifie

$$|G(x) - G(y)| \leq \beta|x - y|. \quad (11)$$

Théorème 13 (Descente de gradient). Soit $J : C^1(\mathbb{R}, \mathbb{R})$ convexe, coercive et telle que J' est β -Lipschitz continue. L'algorithme de descente de gradient de pas $\tau < \frac{2}{\beta}$

$$x_{k+1} = x_k - \tau J'(x_k) \quad (12)$$

converge vers un minimum de J .

Exercice 10 Décrire cet algorithme dans le cas où la fonction à minimiser est

$$f(x) = x^2. \quad (13)$$

4.4 Références

Chloé-Agathe Azencott, "Introduction au Machine Learning", https://cazencott.info/dotclear/public/lectures/IntroML_Azencott.pdf

5 Application : analyse des batteries Li-Ion

Pour la mise en application, nous considérons les signaux que constituent des mesures faites sur des batteries Li-Ion qui proviennent de la référence suivante :

B. SAHA, "Battery Data Set," *NASA AMES Prognostics Data Repository*, 2007

Les signaux sont accessibles à ce lien

<https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>.

et sont décrits comme suit

"Experiments on Li-Ion batteries. Charging and discharging at different temperatures. Records the impedance as the damage criterion. The data set was provided by the NASA Prognostics Center of Excellence (PCoE)."

Le sujet de TP est détaillé dans un autre document.

6 Hors programme

Ci-dessous on présente la notion d'exhaustivité d'une statistique. Cela peut être utile pour vérifier que la variable de synthèse d'un algorithme en ligne, que nous avons noté Y_k , est correctement modélisée, i.e. vérifier qu'elle contient suffisamment d'information.

6.1 Exhaustivité

On s'intéresse ici à la statistique T appliqué à un échantillon $\mathbf{X} = (X_1, X_2, \dots, X_n)$ d'une loi P_θ , où θ représente les paramètres de la loi (ou éventuellement une partie). Nous allons d'abord définir la notions de statistique suffisante (aussi statistique exhaustive selon les auteurs). L'objectif est de comprendre que si T est suffisamment informative, alors on doit pouvoir remplacer \mathbf{X} par $T(\mathbf{X})$. Comment savoir si $T(\mathbf{X})$ engendre une perte d'information par rapport à \mathbf{X} ?

L'intuition est que si T est suffisamment information par rapport à la loi P_θ , alors \mathbf{X} sachant $T(\mathbf{X}) = t$ ne dépend pas de P_θ . Plus formellement, on a la définition suivante.

Définition 14 (Statistique suffisante). La statistique T est dite statistique suffisante pour θ si

$$P_\theta(\mathbf{X}|T(\mathbf{X}) = t) \quad \text{ne dépend pas de } \theta. \quad (14)$$

Autrement dit, si la loi conditionnelle de \mathbf{X} sachant $T(\mathbf{X}) = t$ n'est pas une fonction du paramètre θ .

Exercice 11 Soient $X_1, \dots, X_n \sim \mathcal{B}(\delta)$. Montrer que la statistique suivante est suffisante par rapport à δ

$$T = \sum_{j=1}^n X_j. \quad (15)$$

Référence

V. Monbet, "Statistiques Master Statistique et économétrie Notes de cours", https://perso.univ-rennes1.fr/valerie.monbet/Cours_Stat/cours2010.pdf

7 Éléments de base et rappels

On trouvera dans cette section des éléments sur l'intégration, la transformée de Fourier et le temps-fréquence. Elle a pour objectif de fournir certains prérequis et de faire un rappel, mais son contenu n'est pas l'objet d'intérêt pour l'examen et ne constitue un document de cours.

Remarque 15. Dans cette section, la mesure μ est la mesure de Lebesgue.

7.1 Éléments généraux

Soit $f : \Omega \rightarrow \mathbb{R}$, son support est défini par

$$\text{supp}(f) := \overline{\{x \in \Omega : f(x) \neq 0\}}, \quad (16)$$

où \bar{A} est l'adhérence de l'ensemble A .

7.2 Probabilités

Soit X une variable aléatoire définie sur l'espace probabilisé (Ω, \mathcal{A}, P) . Les éléments de l'espace peuvent se décrire comme :

- l'univers Ω est l'ensemble des résultats possibles ;
- la tribu \mathcal{A} est un ensemble qui contient des éléments dont on peut évaluer la probabilité ;
- P est la mesure de probabilité qui vérifie $P(\Omega) = 1$.

La probabilité que la variable aléatoire $X : \Omega \rightarrow E$ soit contenue dans $S \subseteq E$ s'exprime comme suit

$$P(X \in S) = P(\{\omega \in \Omega | X(\omega) \in S\}) \quad (17)$$

Soient A, B deux variables aléatoires définies sur (Ω, \mathcal{A}, P) .

Définition 16 (Probabilité conditionnelle).

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (18)$$

Théorème 17 (Bayes).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (19)$$

7.3 Intégration

Pour alléger les expressions, $d\mu$ peut être utilisé dans l'intégrale

$$\int_{\mathbb{R}} g d\mu = \int_{\mathbb{R}} g(t) dt. \quad (20)$$

Définition 18. Une fonction f est un élément de $L^p(\mathbb{R})$ si

$$\int_{\mathbb{R}} |f|^p d\mu < \infty. \quad (21)$$

Définition 19. Soient $p \in \mathbb{N}, p > 0$ et $f \in L^p(\mathbb{R})$, la norme p de f est définie par

$$\|f\|_p = \left(\int_{\mathbb{R}} |f|^p d\mu \right)^{\frac{1}{p}}. \quad (22)$$

7.4 Représentations temps-fréquence

Tout d'abord, rappelons la définition de la transformée de Fourier

Définition 20. Soit $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, la transformée de Fourier est définie comme suit

$$\mathcal{F}(f)(\eta) = \int_{\mathbb{R}} f(t) e^{-2i\pi t\eta} dt. \quad (23)$$

Définition 21. Rappelons que pour $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ où f est la fonction (ou signal) d'intérêt et g la fenêtre d'analyse, la TFCT est définie via

$$V_f^g(t, \eta) = \int_{\mathbb{R}} f(\tau) g(t - \tau) e^{-2i\pi\tau\eta} d\tau. \quad (24)$$

Cette transformée est inversible, et le signal peut être obtenu de plusieurs manières.

Proposition 22. La TFCT peut s'inverser si $g(0) \neq 0$

$$f(t) = \frac{1}{g(0)} \int_{\mathbb{R}} V_f^g(t, \eta) d\eta. \quad (25)$$

Si $\int_{\mathbb{R}} g d\mu > 0$

$$\frac{1}{\int_{\mathbb{R}} g d\mu} \int_{\mathbb{R}^2} V_f^g(u, \eta) e^{2i\pi(t-u)\eta}. \quad (26)$$

Si $\mu(\text{supp}(g)) > 0$

$$f(t) = \frac{1}{\|g\|_2} \int_{\mathbb{R}^2} V_f^g(u, \eta) g(t - u) e^{2i\pi(t-u)\eta}. \quad (27)$$

La définition de la TFCT pour les signaux et filtres discrets de taille finie (c'est ce que l'on calcul en pratique) est donnée ci-dessous.

Définition 23. Soient $f \in \mathbb{R}^N, g \in \mathbb{R}^M$

$$V_f^g[m, k] = \sum_{n=1}^N f[n]g[n-m]e^{-2i\pi(n-m)k/N} \quad (28)$$

Elle est inversible elle aussi, mais les formules sont omises.