VQE et QAOA

Léo Durand-Köllner

Février 2021

1 Introduction

Les algorithmes de Shor et de Grover sont prometteurs, mais les ordinateurs quantiques d'aujourd'hui ne permettent pas de les exploiter à leur plein potentiel. Nous sommes encore dans l'ère NISQ (Noisy Intermediate Scale Quantum), et donc on se tourne peu à peu vers une autre catégorie de problèmes qui peuvent être résolus avec le matériel d'aujourd'hui : les algorithmes variationnels.

L'idée est la suivante : un ordinateur classique paramétrise un circuit quantique. En quelques sortes, on caractérise l'état quantique du circuit par des paramètres classiques, et l'on utilise le circuit quantique comme un outil de calcul dirigé par un ordinateur classique. On voit maintenant notre circuit comme une boîte noire : c'est simplement une fonction avec des paramètres à optimiser. On utilise ensuite sur cette fonction des méthodes d'optimisation classique.

En paramétrisant notre circuit quantique, on perd certains états, car on réduit notre gros espace de Hilbert naturel, de dimension 2^n, à un espace plus petit, que l'on appelle *ansatz*. On ne pourrait pas analyser efficacement cet *ansatz* avec un ordinateur classique, et c'est justement là où on tire parti des ordinateurs quantiques. On parle d'algorithme hybride : on utilise à la fois un CPU (Central Processing Unit) et un QPU (Quantum Processing Unit). A chaque fois que l'on a besoin d'un calcul difficile, on utilise le QPU. Mais le contrôle de l'algorithme est laissé au CPU.

2 VQE : Variational Quantum Eigensolver

On se donne un hamiltonien $\mathcal{H}$ qui n'est rien d'autre qu'une matrice hermitienne. Un hamiltonien représente l'énergie d'un système quantique : les énergies accessibles sont les valeurs propres de $\mathcal{H}$. On cherche le niveau d'énergie minimal du système, *i.e.* la valeur propre minimale de notre matrice hermitienne. VQE nous permet justement de faire cela.

Comme expliqué dans l'introduction, l'idée d'un algorithme variationnel (VQE en est un) est d'avoir un circuit quantique qui prépare des états à la demande. Plus précisément, ce circuit produit un état quantique $|\psi(\theta_1, ..., \theta_m)\rangle$ paramétrisé par des angles θ_i, et ces angles sont donnés par un optimisateur classique.

On définit la fonction f qui, à un ensemble de paramètres $\{\theta_i\}$, donne le niveau d'énergie de $|\psi(\theta_1, ..., \theta_m)\rangle$. Formellement, on cherche à minimiser $f : (\theta_1, ..., \theta_m) \mapsto \langle\psi(\theta_1, ..., \theta_m)|\mathcal{H}|\psi(\theta_1, ..., \theta_m)\rangle$. Le calcul de f est donc fait par le QPU. L'algorithme est plutôt simple :

1. Initialisation : On démarre par un état $|\psi(\theta_1^0, ..., \theta_m^0)\rangle$ (aléatoire), et on calcule $f(\theta_1^0, ..., \theta_m^0)$ l'énergie associée.
2. On utilise une méthode d'optimisation numérique (*e.g* descente du gradient) pour déterminer de nouveaux paramètres $(\theta_1^{i+1}, ..., \theta_m^{i+1})$ à partir des paramètres précédents $(\theta_1^i, ..., \theta_m^i)$.
3. On évalue $f(\theta_1^{i+1}, ..., \theta_m^{i+1})$. Puis goto 2.

On voit que l'on peut faire une analogie avec le machine learning. Pour un réseau de neurones, on dispose d'une fonction de coût qui dépend des biais et des poids, et on cherche à trouver les bonnes valeurs de paramètres pour minimiser l'erreur en itérant sur une correction des paramètres d'entrée. C'est exactement ce que l'on fait ici !

Maintenant, reste encore 3 questions à éclairer :
— Comment fabriquer l'état $|\psi(\theta_1, ..., \theta_m)\rangle$ pour des angles arbitraires ?
— Pourquoi converge-t-on vers un vecteur propre de valeur propre minimale ?
— Comment mesurer $\langle\psi(\theta_1, ..., \theta_m)|\mathcal{H}|\psi(\theta_1, ..., \theta_m)\rangle$?

Pour la première question, cela dépend évidemment de la fonction étudiée, et on en reparle avec QAOA. Le principe variationnel répond au deuxième point. Ce principe affirme que pour tout opérateur hermitien $\mathcal{H}$ qui admet une valeur propre minimale λ_{min}, et pour tout état quantique $|\psi\rangle$, $\langle\psi|\mathcal{H}|\psi\rangle \geq \lambda_{min}$. Par conséquent si on finit par trouver le minimum de f, on sait donc que l'état quantique associé est un vecteur propre de $\mathcal{H}$, et pas autre chose !

La démonstration est assez simple : on se donne un état quelconque $|\psi\rangle$, et l'état $|\psi_{min}\rangle$ associé à la valeur propre minimale. On peut donc écrire, de façon unique, $|\psi\rangle = |\psi_{min}\rangle + |\psi_{min}^\perp\rangle$, où $|\psi_{min}^\perp\rangle$ est orthogonal à $|\psi_{min}\rangle$. Ainsi :

$$\begin{aligned}\exists\alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1,& \\ \langle\psi|\mathcal{H}|\psi\rangle &= (\bar{\alpha}\langle\psi_{min}| + \bar{\beta}\langle\psi_{min}^\perp|)\mathcal{H}(\alpha|\psi_{min}\rangle + \beta|\psi_{min}^\perp\rangle) \\ &= |\alpha|^2\langle\psi_{min}|\mathcal{H}|\psi_{min}\rangle + \bar{\alpha}\beta\langle\psi_{min}|\mathcal{H}|\psi_{min}^\perp\rangle \\ &+ \bar{\beta}\alpha\langle\psi_{min}^\perp|\mathcal{H}|\psi_{min}\rangle + |\beta|^2\langle\psi_{min}^\perp|\mathcal{H}|\psi_{min}^\perp\rangle\end{aligned}$$

On a $\langle\psi_{min}|\mathcal{H}|\psi_{min}\rangle = \lambda_{min}$. Pour les autres termes, le théorème spectral nous donne l'existence d'une base orthonormale formée de vecteurs propres de $\mathcal{H}$. Par conséquent $|\psi_{min}^\perp\rangle$ s'écrit comme une combinaison linéaire de vecteurs propres qui ne sont pas $|\psi_{min}\rangle$. Donc d'une part $\mathcal{H}|\psi_{min}^\perp\rangle \in \{\psi_{min}\}^\perp$, et d'autre part $\langle\psi_{min}^\perp|\mathcal{H}|\psi_{min}^\perp\rangle \geq \lambda_{min}$. Ainsi :

$$\langle\psi|\mathcal{H}|\psi\rangle \geq (|\alpha|^2 + |\beta|^2)\lambda_{min} = \lambda_{min}$$

Maintenant, concernant le dernier point, on utilie un théorème qui nous permet d'affirmer qu'un opérateur hermitien peut toujours se décomposer sous la forme :

$$\mathcal{H} = \sum_{i,\alpha} h_i^\alpha \sigma_i^\alpha + \sum_{i,j,\alpha,\beta} h_{i,j}^{\alpha,\beta} \sigma_i^\alpha \sigma_j^\beta + ...$$

où les h sont des réels, les lettres latines sont juste des indices, et les lettres grecques valent soit X, Y ou Z, de sorte que σ_i^α représente une porte de Pauli α agissant sur le qbit i. Dans la suite, on considère des hamiltoniens qui ont un nombre de termes polynomial en la taille du système étudié. Dans la suite, on suppose connaître cette décomposition.

Par conséquent, si on se donne un état $|\psi\rangle$, évaluer $\langle\psi|\mathcal{H}|\psi\rangle$ revient à évaluer les $\langle\psi|\sigma_i^\alpha|\psi\rangle$, puis les $\langle\psi|\sigma_i^\alpha\sigma_j^\alpha|\psi\rangle$, etc... On est donc ramenés à des portes de Pauli élémentaires ! Pour déterminer ces espérances, on ne dispose que des probabilités $P(0)$ et $P(1)$.

Pour la matrice Z :
On commence par celle-là parce que c'est la plus simple. En effet, on peut remarquer que $\mathbf{Z} = |0\rangle\langle 0| - |1\rangle\langle 1|$, et donc que $\mathbf{Z}|\psi\rangle = \langle 0|\psi\rangle|0\rangle - \langle 1|\psi\rangle|1\rangle$. D'où $\langle\psi|\mathbf{Z}|\psi\rangle = \langle 0|\psi\rangle\langle\psi|0\rangle - \langle 1|\psi\rangle\langle\psi|1\rangle = |\langle 0|\psi\rangle|^2 - |\langle 1|\psi\rangle|^2 = P(0) - P(1)$.

Pour la matrice X :
On sait que $\mathbf{X} = \mathbf{HZH}$ donc $\langle\psi|\mathbf{X}|\psi\rangle = (\langle\psi|\mathbf{H})\mathbf{Z}(\mathbf{H}|\psi\rangle)$, et comme $\mathbf{H}$ est hermitienne, $\langle\psi|\mathbf{X}|\psi\rangle = \langle\phi|\mathbf{Z}|\phi\rangle$ avec $|\phi\rangle = \mathbf{H}|\psi\rangle$.

Pour la matrice Y :
On sait que $\mathbf{Y} = \mathbf{S}^\dagger\mathbf{XS}$, donc $\langle\psi|\mathbf{Y}|\psi\rangle = (\langle\psi|\mathbf{S}^\dagger\mathbf{H})\mathbf{Z}(\mathbf{HS}|\psi\rangle) = \langle\phi|\mathbf{Z}|\phi\rangle$ avec $|\phi\rangle = \mathbf{HS}|\psi\rangle$.

Pour ce qui est des produits de matrices de Pauli, il suffit de généraliser. Par exemple, si on regarde $\mathbf{Z}\otimes\mathbf{X}$, et bien $\mathbf{Z}\otimes\mathbf{X} = \mathbf{H}_0(\mathbf{Z}\otimes\mathbf{Z})\mathbf{H}_0$, et donc $\langle\psi|\mathbf{Z}\otimes\mathbf{X}|\psi\rangle = (\langle\psi|\mathbf{H}_0)(\mathbf{Z}\otimes\mathbf{Z})(\mathbf{H}_0|\psi\rangle)$, et donc on se ramène à l'évaluation de l'espérance pour $\mathbf{Z}\otimes\mathbf{Z} = |0\rangle\langle 0| - |1\rangle\langle 1| + |2\rangle\langle 2| - |3\rangle\langle 3|$, qui vaut donc simplement $\langle\psi|\mathbf{Z}\otimes\mathbf{Z}|\psi\rangle = P(0) - P(1) + P(2) - P(3)$.

3 Des problèmes d'optimisation combinatoire aux opérateurs quantiques

Un problème d'optimisation combinatoire est défini par la donnée d'une fonction (appelée *fonction objectif*) $f : \{0,1\}^n \mapsto \mathbb{R}$, *i.e.* une fonction qui associe à une suite de valeurs booléennes (bits) une certaine valeur. Le but du problème est de trouver une séquence de bits $s_1...s_n$ qui maximise f. Pour cela, on caractérise f par un hamiltonien, que l'on note $\mathbf{C}$, défini par :

$$\mathbf{C}|s\rangle = f(s)|s\rangle$$

On définit ainsi un opérateur diagonal dont les valeurs propres sont les valeurs $f(z)$, et les vecteurs propres associés sont les états de la base canonique, c'est-à-dire :

$$\begin{aligned}\mathbf{C} &= diag(f(0), f(1), ..., f(2^n-1)) \\ &= \begin{pmatrix} f(0) & 0 & ... & 0 \\ 0 & f(1) & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & f(2^n-1) \end{pmatrix}\end{aligned}$$

On peut alors reformuler le problème de la manière suivante : trouver un état associé au niveau d'énergie le plus haut de cet hamiltonien, c'est-à-dire

un vecteur propre de $\mathbf{C}$ associé à la valeur propre la plus élevée.

Prenons l'exemple du problème Max-Cut. On se donne un graphe $G = (V, E)$, et on cherche une coupe qui soit traversée par un maximum d'arêtes. Formalisons ce problème : on pose $V = V_1 \cup V_2$ avec $V_1 \cap V_2 = \varnothing$ la coupe choisie, et $\forall i \in V, s_i = 0$ si $i \in V_1$, 1 sinon. On note $\forall i \in V, z_i = (-1)^{s_i}$. On définit alors la fonction objectif qui compte le nombre d'arêtes qui traversent la coupe :

$$\forall s \in \{0,1\}^n, f(s) = \frac{1}{2} \sum_{\{i,j\} \in E} (1 - z_i z_j)$$

Bien entendu, l'opérateur $\mathbf{C}$ associé à f est bien trop gros pour être directement implémenté, puisqu'il est de taille exponentielle en n. Mais on peut considérablement réduire son expression dans le cas du problème Max-Cut. En effet, on peut écrire :

$$\mathbf{C} = \frac{1}{2} \sum_{\{i,j\} \in E} I - \mathbf{Z}_i \mathbf{Z}_j$$

où $\mathbf{Z}$ est la porte Z de Pauli. Cet opérateur est de suite beaucoup plus simple à implémenter, puisque l'on utilise seulement $\mathcal{O}(n^2)$ portes de Pauli.

4 QAOA : Quantum Approximate Optimization Algorithm

On rappelle que l'on cherche à trouver un vecteur propre de $\mathbf{C}$ associé à la valeur propre maximale, ou autrement dit un vecteur propre de $-\mathbf{C}$ associé à la valeur propre minimale. Cela fait-il penser à quelque chose ? Oui tout à fait, on utilise VQE ! Pour cela, QAOA paramétrise l'état $|\psi\rangle$ par deux angles, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ et $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p)$, où p désigne un entier fixé.

4.1 L'ansatz $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$

On revient ici sur le choix et l'implémentation de l'état paramétrisé $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$. On suppose disposer des portes usuelles $\mathbf{H}$, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{R}_z(\theta) = e^{-i\frac{\theta}{2}\mathbf{Z}}$, et $\mathbf{C}_{10}$ (CNOT). Sans faire durer plus longtemps le suspense, voici l'expression de cet état :

$$|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = e^{-i\beta_p \mathbf{B}} e^{-i\gamma_p \mathbf{C}} ... e^{-i\beta_1 \mathbf{B}} e^{-i\gamma_1 \mathbf{C}} \mathbf{H}^{\otimes n} |0\rangle$$

où $\mathbf{B} = \sum_{i \in V} \mathbf{X}_i$.

On cherche donc à implémenter les opérateurs $\mathbf{U}(\beta, \mathbf{B}) = e^{-i\beta\mathbf{B}}$ et $\mathbf{U}(\gamma, \mathbf{C}) = e^{-i\gamma\mathbf{C}}$, pour deux réels β et γ. On commence par $\mathbf{U}(\gamma, \mathbf{C})$, dont l'étude se réduit à $e^{i\frac{\gamma}{2}\mathbf{Z}_i\mathbf{Z}_j}$. En effet, puisque les exponentielles de matrices commutent,

$$\mathbf{U}(\gamma, \mathbf{C}) = \prod_{\{i,j\} \in E} e^{-i\frac{\gamma}{2}(\mathbf{I} - \mathbf{Z}_i\mathbf{Z}_j)} = e^{-i\frac{\gamma}{2} \times Card(E)} \prod_{\{i,j\} \in E} e^{i\frac{\gamma}{2}\mathbf{Z}_i\mathbf{Z}_j}$$

On ne se soucie pas de la phase globale dans la suite car elle n'influence pas la mesure au final. Maintenant, dans le cas n=2, comme $\frac{1}{2}\mathbf{ZZ}$ admet les vecteurs de la base canonique comme vecteurs propres associés aux valeurs propres -1 et 1, on a :

1. $e^{i\frac{\gamma}{2}\mathbf{ZZ}}|00\rangle = e^{i\frac{\gamma}{2}}|00\rangle$
2. $e^{i\frac{\gamma}{2}\mathbf{ZZ}}|01\rangle = e^{-i\frac{\gamma}{2}}|01\rangle$
3. $e^{i\frac{\gamma}{2}\mathbf{ZZ}}|10\rangle = e^{-i\frac{\gamma}{2}}|10\rangle$

4. $e^{i\frac{\gamma}{2}\mathbf{ZZ}}|11\rangle = e^{i\frac{\gamma}{2}}|11\rangle$

On reconnait donc que $e^{i\frac{\gamma}{2}\mathbf{ZZ}}|ab\rangle = e^{i\frac{\gamma}{2}(-1)^{a+b}}|ab\rangle$. Mais on sait que $\mathbf{C}_{10}|ab\rangle = |a\rangle|a \oplus b\rangle$ et que $\mathbf{R}_Z(\theta)|a\rangle = e^{-i(-1)^a\frac{\theta}{2}}$. On en déduit donc que $\mathbf{R}_{Z,0}(-\gamma)\mathbf{C}_{10}|ab\rangle = e^{i\frac{\gamma}{2}(-1)^{a\oplus b}}|a\rangle|a \oplus b\rangle$. D'où :

$$\begin{aligned}\mathbf{C}_{10}\mathbf{R}_{Z,0}(-\gamma)\mathbf{C}_{10}|ab\rangle &= e^{i\frac{\gamma}{2}(-1)^{a+b}}|ab\rangle \\ &= e^{i\frac{\gamma}{2}\mathbf{ZZ}}|ab\rangle\end{aligned}$$

Cette relation est vraie sur la base canonique, donc partout. Ainsi on obtient l'expression finale (en omettant la phase globale) :

$$\mathbf{U}(\gamma, \mathbf{C}) = \prod_{\{i,j\}\in E} \mathbf{C}_{i,j}\mathbf{R}_{Z,i}(-\gamma)\mathbf{C}_{i,j}$$

Concernant $\mathbf{U}(\beta, \mathbf{B})$, on remarque que $\mathbf{X} = \mathbf{HZH}$, ce qui nous donne, en utilisant le développement en série de Taylor de l'exponentielle, que $\forall a \in \mathbb{C}, e^{a\mathbf{X}} = \mathbf{H}e^{a\mathbf{Z}}\mathbf{H}$:

$$\begin{aligned}\mathbf{U}(\beta, \mathbf{B}) &= \prod_{i\in V} e^{-i\frac{\beta}{2}\mathbf{X}_i} \\ &= \prod_{i\in V} \mathbf{H}_i e^{-i\frac{\beta}{2}\mathbf{Z}_i}\mathbf{H}_i \\ &= \prod_{i\in V} \mathbf{H}_i\mathbf{R}_{Z,i}(\beta)\mathbf{H}_i\end{aligned}$$

4.2 Calcul de l'espérance

Dans cette section, on répond à la question : comment calculer $\langle \boldsymbol{\gamma}, \boldsymbol{\beta}|\mathbf{C}|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$?

Pour cela, on pourrait essayer de retrouver la décomposition en sommes d'opérateurs élémentaires de Pauli (cf section VQE), mais ici on va même trouver plus simple. En effet, on peut directement voir que pour un état quelconque $|\psi\rangle$,

$$\langle\psi|\mathbf{C}|\psi\rangle = \frac{1}{2}\sum_{\{i,j\}\in E} 1 - \langle\psi|\mathbf{Z}_i\mathbf{Z}_j|\psi\rangle$$

Or,

$$\begin{aligned}\mathbf{Z}_1\mathbf{Z}_0 &= diag(1,1,-1,-1)diag(1,-1,1,-1) \\ &= diag(1,-1,-1,1) \\ &= |00\rangle\langle 00| - |01\rangle\langle 01| - |10\rangle\langle 10| + |11\rangle\langle 11|\end{aligned}$$

donc on en déduit que $\langle\psi|\mathbf{Z}_1\mathbf{Z}_0|\psi\rangle = P(00) - P(01) - P(10) + P(11)$. En généralisant, on obtient :

$$\begin{aligned}\langle\psi|\mathbf{Z}_i\mathbf{Z}_j|\psi\rangle = P(|\psi\rangle \wedge (2^i + 2^j) = 0) - P(|\psi\rangle \wedge (2^i + 2^j) = 2^i) \\ - P(|\psi\rangle \wedge (2^i + 2^j) = 2^j) + P(|\psi\rangle \wedge (2^i + 2^j) = 2^i + 2^j)\end{aligned}$$

où le $\wedge$ désigne l'opération ET logique bit à bit.