بهینه سازی محدب ۱

نیم‌سال اول ۱۴۰۳-۱۴۰۴

دکتر امیری

دانشکده مهندسی برق

گزارش پروژه

محمد مهدی رزمجو-۴۰۰۱۰۱۲۷۲ ابوالفضل گرگوری مطلق-۴۰۰۱۰۱۸۴۸

مقدمه

الگوریتم‌های کوانتومی در سال‌های اخیر به عنوان ابزارهای قدرتمندی برای حل مسائل بهینه‌سازی ترکیبی مطرح شده‌اند. یکی از این مسائل، **مسئله‌ی بیشینه‌برش (MaxCut)** است که در زمینه‌هایی مانند شبکه‌های ارتباطی، علوم داده، و فیزیک آماری کاربرد گسترده‌ای دارد. در این پروژه، یک الگوریتم کوانتومی ارائه شده است که راه‌حل‌های تقریبی برای مسائل بهینه‌سازی ترکیبی تولید می‌کند. عملکرد این الگوریتم به یک عدد صحیح $p \geq 1$ وابسته بوده و با افزایش مقدار p، دقت تقریب بهبود می‌یابد. مدار کوانتومی مورد استفاده از گیت‌های یکتایی تشکیل شده است که بیشینه‌ی محلی‌بودن آن‌ها متناسب با بیشینه‌ی محلی‌بودن تابع هدف است. عمق این مدار نیز متناسب با مقدار p و در بدترین حالت، برابر با تعداد قیود افزایش می‌یابد.

در صورتی که p مقدار ثابتی داشته باشد و مستقل از اندازه‌ی ورودی باشد، الگوریتم از یک پیش‌پردازش کلاسیک کارآمد بهره می‌برد. اما زمانی که p با افزایش اندازه‌ی ورودی رشد کند، رویکرد متفاوتی برای اجرا در نظر گرفته می‌شود.

معرفی مسئله

مسائل بهینه‌سازی ترکیبیاتی توسط n بیت و m بند مشخص می‌شوند. هر بند یک قید روی زیرمجموعه‌ای از بیت‌هاست که برای برخی انتساب‌های آن بیت‌ها ارضا شده و برای انتساب‌های دیگر ارضا نمی‌شود. تابع هدف که روی رشته‌های n بیتی تعریف می‌شود، تعداد بندهای ارضا شده است:

$$C(z) = \sum_{\alpha=1}^{m} C_\alpha(z)$$

که در آن $z = z_1 z_2 \dots z_n$ یک رشته بیتی بوده و $C_\alpha(z) = 1$ اگر z بند α را ارضا کند و در غیر این صورت 0 است. معمولاً C_α فقط به تعداد کمی از بیت‌ها وابسته است. مسئله SAT (رضایت‌پذیری) می‌پرسد آیا رشته‌ای وجود دارد که همه بندها را ارضا کند. MaxSat به دنبال رشته‌ای است که تابع هدف را بیشینه کند. بهینه‌سازی تقریبی به دنبال رشته‌ای z است که $C(z)$ به بیشینه C نزدیک باشد. در این پروژه از یک الگوریتم کوانتومی کلی برای بهینه‌سازی تقریبی استفاده می کنیم. عملکرد آن را در موارد خاص MaxCut بررسی می کنیم.

کامپیوتر کوانتومی در فضای هیلبرت 2^n بعدی با بردارهای پایه محاسباتی $|z\rangle$ کار می‌کند و ما $C(z)$ را به عنوان عملگری که در پایه محاسباتی قطری است در نظر می‌گیریم. یک عملگر یکانی $U(C,\gamma)$ را تعریف می‌کنیم که به زاویه γ وابسته است:

$$U(C,\gamma) = e^{-i\gamma C} = \prod_{\alpha=1}^{m} e^{-i\gamma C_\alpha}$$

تمام جمله‌های این حاصلضرب جابه‌جاشدنی هستند زیرا در پایه محاسباتی قطری بوده و محلّیت هر جمله برابر با محلّیت بند α است. از آنجا که C مقادیر ویژه صحیح دارد، می‌توان γ را بین 0 تا 2π محدود کرد. عملگر B را به عنوان مجموع تمام عملگرهای σ^x تک بیتی تعریف می‌کنیم:

$$B = \sum_{j=1}^{n} \sigma_j^x$$

حال حاصلضرب وابسته به β از عملگرهای تک بیتی جابه‌جاشدنی را تعریف می‌کنیم:

$$U(B,\beta) = e^{-i\beta B} = \prod_{j=1}^{n} e^{-i\beta\sigma_j^x}$$

که β از 0 تا π تغییر می‌کند. حالت اولیه $|s\rangle$ برهم‌نهی یکنواخت بر روی حالت‌های پایه محاسباتی خواهد بود:

$$|s\rangle = \frac{1}{\sqrt{2^n}} \sum_z |z\rangle$$

برای هر عدد صحیح $p \geq 1$ و $2p$ زاویه $\gamma_1 \dots \gamma_p \equiv \boldsymbol{\gamma}$ و $\beta_1 \dots \beta_p \equiv \boldsymbol{\beta}$، حالت کوانتومی وابسته به زاویه را به صورت زیر تعریف می‌کنیم:

$$|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle = U\left(B,\beta_p\right) U\left(C,\gamma_p\right) \cdots U\left(B,\beta_1\right) U\left(C,\gamma_1\right) |s\rangle.$$

حتی بدون استفاده از ساختار نمونه مسئله، این حالت را می‌توان با مداری کوانتومی با عمق حداکثر $mp + p$ تولید کرد. فرض کنید F_p مقدار انتظاری C در این حالت باشد:

$$F_p(\boldsymbol{\gamma},\boldsymbol{\beta}) = \langle \boldsymbol{\gamma},\boldsymbol{\beta}|C|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle.$$

و M_p بیشینه F_p نسبت به زاویه‌ها باشد:

$$M_p = \max_{\boldsymbol{\gamma},\boldsymbol{\beta}} F_p(\boldsymbol{\gamma},\boldsymbol{\beta}).$$

توجه کنید که بیشینه‌سازی در $p-1$ را می‌توان به عنوان یک بیشینه‌سازی مقید در p در نظر گرفت، بنابراین:

$$M_p \geq M_{p-1}.$$

علاوه بر این، بعداً نشان خواهیم داد که:

$$\lim_{p\to\infty} M_p = \max_z C(z).$$

این نتایج روشی برای طراحی الگوریتم پیشنهاد می‌دهند. یک p انتخاب می کنیم و با مجموعه‌ای از زاویه‌های $(\boldsymbol{\gamma},\boldsymbol{\beta})$ شروع می کنیم که به نحوی F_p را تا حد امکان بزرگ می‌کنند. بایستی از کامپیوتر کوانتومی برای تولید حالت $|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle$ استفاده کرد. با اندازه‌گیری در پایه محاسباتی، یک رشته z به دست می آوریم و $C(z)$ را محاسبه می کنیم. این فرآیند را با همان زاویه‌ها تکرار می کنیم. تکرار کافی، رشته z را با $C(z)$ بسیار نزدیک یا بزرگتر از $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ تولید خواهد کرد. مشکل این است که از قبل مشخص نیست چگونه زاویه‌های مناسب را انتخاب کنیم.

اگر p با n رشد نکند، یک امکان این است که کامپیوتر کوانتومی را با زاویه‌های $(\boldsymbol{\gamma},\boldsymbol{\beta})$ انتخاب‌شده از یک شبکه ریز در مجموعه فشرده $[0,2\pi]^p \times [0,\pi]^p$ اجرا کرده و در شبکه جستجو می کنیم تا بیشینه F_p را بیابیم. از آنجا که مشتقات جزئی $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ در رابطه با $\mathcal{O}\left(m^2+mn\right)$ محدود می‌شوند، این جستجو به صورت کارآمد رشته z را تولید می‌کند که $C(z)$ نزدیک به M_p یا بزرگتر از آن است. با این حال، در بخش بعدی نشان می‌دهیم که اگر p با n رشد نکند و هر بیت در حداکثر تعداد ثابتی از بندها دخیل باشد، یک محاسبه کلاسیک کارآمد وجود دارد که زاویه‌های بیشینه‌کننده F_p را تعیین می‌کند. سپس از این زاویه‌ها برای اجرای کامپیوتر کوانتومی و تولید حالت $|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle$ استفاده می‌شود که در پایه محاسباتی اندازه‌گیری شده تا رشته z به دست آید. میانگین $C(z)$ برای رشته‌های حاصل از این روش برابر M_p است.

اکنون توضیح می‌دهیم که چگونه برای p ثابت می‌توانیم پیش‌پردازش کلاسیک انجام داده و زاویه‌های $\boldsymbol{\gamma}$ و $\boldsymbol{\beta}$ را که $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ را بیشینه می‌کنند، تعیین کنیم. این روش به صورت کلی کاربرد دارد، اما آن را برای یک مسئله خاص، یعنی MaxCut برای گراف‌های با درجه محدود، توضیح می‌دهیم. ورودی یک گراف با n رأس و مجموعه یال‌های $\{\langle jk\rangle\}$ به اندازه m است. هدف یافتن رشته z است که عبارت زیر را بیشینه کند:

$$C = \sum_{\langle jk\rangle} C_{\langle jk\rangle}$$

که در آن:

$$C_{\langle jk\rangle} = \frac{1}{2}\left(-\sigma_j^z\sigma_k^z + 1\right)$$

حالا داریم:

$$F_p(\boldsymbol{\gamma},\boldsymbol{\beta}) = \sum_{\langle jk\rangle} \langle s| U^\dagger\left(C,\gamma_1\right)\cdots U^\dagger\left(B,\beta_p\right) C_{\langle jk\rangle} U\left(B,\beta_p\right)\cdots U\left(C,\gamma_1\right)|s\rangle$$

عملگر مرتبط با یال $\langle jk\rangle$ را در نظر بگیرید:

$$U^\dagger\left(C,\gamma_1\right)\cdots U^\dagger\left(B,\beta_p\right) C_{\langle jk\rangle} U\left(B,\beta_p\right)\cdots U\left(C,\gamma_1\right)$$

این عملگر تنها شامل کیوبیت‌های j و k و کیوبیت‌هایی است که فاصله آن‌ها روی گراف از j یا k کمتر یا برابر p باشد. برای درک این موضوع، حالت $p=1$ را در نظر بگیرید که عبارت فوق به شکل زیر است:

$$U^\dagger\left(C,\gamma_1\right) U^\dagger\left(B,\beta_1\right) C_{\langle jk\rangle} U\left(B,\beta_1\right) U\left(C,\gamma_1\right)$$

عوامل موجود در عملگر $U\left(B,\beta_1\right)$ که شامل کیوبیت‌های j یا k نیستند، از $C_{\langle jk\rangle}$ عبور کرده و حذف می‌شوند. بنابراین داریم:

$$U^\dagger\left(C,\gamma_1\right) e^{i\beta_1\left(\sigma_j^x+\sigma_k^x\right)} C_{\langle jk\rangle} e^{-i\beta_1\left(\sigma_j^x+\sigma_k^x\right)} U\left(C,\gamma_1\right)$$

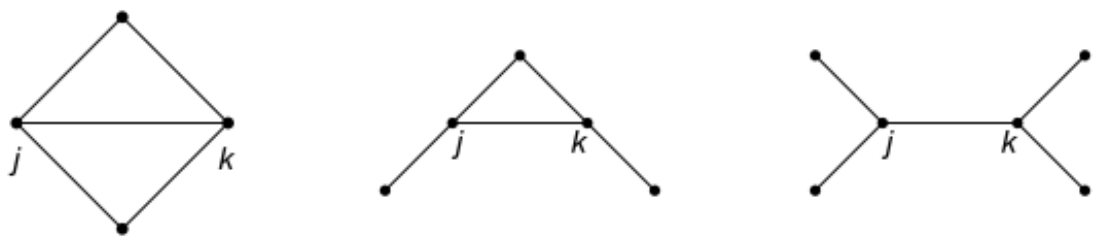
هر عاملی در عملگر $U\left(C,\gamma_1\right)$ که شامل کیوبیت‌های j یا k نباشد، از عبارت عبور کرده و حذف می‌شود. بنابراین عملگر در رابطه بالا $U^\dagger\left(C,\gamma_1\right)\cdots U^\dagger\left(B,\beta_p\right) C_{\langle jk\rangle} U\left(B,\beta_p\right)\cdots U\left(C,\gamma_1\right)$ و یال‌های مجاور آن و کیوبیت‌های روی آن‌هاست. برای هر p، عملگر در رابطه $\langle jk\rangle$ تنها شامل یال‌هایی است که حداکثر p گام از $\langle jk\rangle$ فاصله دارند.

به رابطه $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ بازگردید و توجه کنید که حالت $|s\rangle$ حاصلضرب حالت‌های ویژه σ^x است:

$$|s\rangle = |+\rangle_1|+\rangle_2\ldots|+\rangle_n$$

بنابراین هر جمله در رابطه $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ تنها به زیرگرافی وابسته است که شامل کیوبیت‌های j و k و کیوبیت‌هایی است که در فاصله حداکثر p دوری از آن‌ها قرار دارند. هر یک از این زیرگراف‌ها شامل تعدادی کیوبیت است که مستقل از n است (زیرا درجه گراف محدود است) و این به ما اجازه می‌دهد F_p را بر اساس زیرسیستم‌های کوانتومی با ابعاد مستقل از n محاسبه کنیم.

مثال: به عنوان مثال، مسئله MaxCut را برای گراف‌های با درجه ثابت ۳ در نظر بگیرید. برای $p=1$، تنها زیرگراف‌های ممکن برای یال $\langle jk\rangle$ به صورت زیر هستند:



برای هر زیرگراف G، عملگر C_G را تعریف می‌کنیم که C محدودشده به G است:

$$C_G = \sum_{\langle \ell\ell'\rangle\in G} C_{\langle \ell\ell'\rangle}$$

و عملگر مرتبط با آن:

$$U\left(C_G,\gamma\right) = e^{-i\gamma C_G}$$

همچنین تعریف می‌کنیم:

$$B_G = \sum_{j\in G}\sigma_j^x$$

و

$$U\left(B_G,\beta\right)=e^{-i\beta B_G}$$

حالت $|s,G\rangle$ را به صورت زیر در نظر بگیرید:

$$|s,G\rangle=\prod_{\ell\in G}|+\rangle_\ell$$

به رابطه $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ بازبرمیگردیم. هر یال $\langle j,k\rangle$ در این جمع با یک زیرگراف $g(j,k)$ مرتبط است و سهمی در F_p به اندازه زیر دارد:

$$\langle s,g(j,k)|U^\dagger\left(C_{g(j,k)},\gamma_p\right)\cdots U^\dagger\left(B_{g(j,k)},\beta_1\right)C_{\langle jk\rangle}U\left(B_{g(j,k)},\beta_1\right)\cdots U\left(C_{g(j,k)},\gamma_p\right)|s,g(j,k)\rangle$$

جمع در رابطه $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ بر روی تمام یال‌ها است، اما اگر دو یال $\langle jk\rangle$ و $\langle j'k'\rangle$ زیرگراف‌های هم‌ریختی تولید کنند، آنگاه توابع متناظر از $(\boldsymbol{\gamma},\boldsymbol{\beta})$ یکسان خواهند بود. بنابراین می‌توان جمع در $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ را به عنوان جمعی بر روی انواع زیرگراف‌ها در نظر گرفت.
فرض کنید

$$f_g(\boldsymbol{\gamma},\boldsymbol{\beta})=\langle s,g(j,k)|U^\dagger\left(C_{g(j,k)},\gamma_1\right)\cdots U^\dagger\left(B_{g(j,k)},\beta_p\right)C_{\langle jk\rangle}U\left(B_{g(j,k)},\beta_p\right)\cdots$$
$$U\left(C_{g(j,k)},\gamma_1\right)|s,g(j,k)\rangle$$

که در آن $g(j,k)$ یک زیرگراف از نوع g است. در این صورت F_p به صورت زیر خواهد بود:

$$F_p(\boldsymbol{\gamma},\boldsymbol{\beta})=\sum_g w_g f_g(\boldsymbol{\gamma},\boldsymbol{\beta})$$

که w_g تعداد تکرار زیرگراف g در جمع یال‌های اصلی است. توابع f_g به n و m وابسته نیستند. تنها وابستگی به n و m از طریق وزن‌های w_g است که مستقیماً از گراف اصلی خوانده می‌شوند. توجه کنید که مقدار انتظاری در رابطه $f_g(\boldsymbol{\gamma},\boldsymbol{\beta})$ تنها شامل کیوبیت‌های زیرگراف نوع g ویبت‌هایی که می‌توانند در $\langle s,g(j,k)|U^\dagger\left(C_{g(j,k)},\gamma_p\right)\cdots U^\dagger\left(B_{g(j,k)},\beta_1\right)C_{\langle jk\rangle}U\left(B_{g(j,k)},\beta_1\right)\cdots U\left(C_{g(j,k)},\gamma_p\right)|s,g(j,k)\rangle$ ظاهر شوند، زمانی است که زیرگراف یک درخت باشد. برای گرافی با حداکثر درجه ν، تعداد کیوبیت‌ها در این درخت برابر است با:

$$q_{tree}=2\left[\frac{(\nu-1)^{p+1}-1}{(\nu-1)-1}\right]$$

(یا $2p+2$ اگر $\nu=2$) که مستقل از n و m است. برای هر p، تنها تعداد محدودی از انواع زیرگراف‌ها وجود دارد.
با استفاده از رابطه $f_g(\boldsymbol{\gamma},\boldsymbol{\beta})$، $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ در رابطه $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})=\sum_g w_g f_g(\boldsymbol{\gamma},\boldsymbol{\beta})$ را می‌توان روی یک کامپیوتر کلاسیک با منابعی که با n رشد نمی‌کنند، محاسبه کرد. هر f_g شامل عملگرها و حالت‌ها در فضای هیلبرتی با بعد حداکثر $2^{q_{tree}}$ است. البته برای pهای بزرگ، این ممکن است فراتر از فناوری کلاسیک فعلی باشد، اما نیازمندی‌های منابع با n افزایش نمی‌یابند.
برای اجرای الگوریتم کوانتومی، ابتدا $(\boldsymbol{\gamma},\boldsymbol{\beta})$هایی که F_p را بیشینه می‌کنند، پیدا می‌کنیم. تنها وابستگی به n و m در وزن‌های w_g است که به راحتی محاسبه می‌شوند. پس از یافتن بهترین $(\boldsymbol{\gamma},\boldsymbol{\beta})$، از کامپیوتر کوانتومی برای تولید حالت $|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle$ داده‌شده در رابطه $|\boldsymbol{\gamma},\boldsymbol{\beta}\rangle$ استفاده می‌کنیم. سپس در پایه محاسباتی اندازه‌گیری کرده و رشته z را به دست آورده و $C(z)$ را محاسبه می‌کنیم. با تکرار این فرآیند، نمونه‌ای از مقادیر $C(z)$ بین ۰ تا $+m$ به دست می‌آید که میانگین آن $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$ است. نتیجه‌ای حداقل برابر $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})-1$ با احتمال $1-\frac{1}{m}$ و با $\mathcal{O}(m\log m)$ تکرار به دست خواهد آمد.