

بهینه‌سازی با الگوریتم‌های کلاسیکی و کوانتومی تقویت شده

ابوالفضل رمضان‌پور*

بخش فیزیک، دانشگاه شیراز، شیراز

پست الکترونیکی: aramezanpour@gmail.com

(دریافت مقاله: ۱۴۰۴/۰۹/۰۳؛ دریافت نسخه نهایی: ۱۴۰۴/۰۹/۱۵)

چکیده:

بسیاری از مسئله‌های مهم مانند پیدا کردن حالت تعادلی یک سامانه فیزیکی و یادگیری شبکه‌های عصبی را می‌توان به شکل یک مسئله بهینه‌سازی فرمول‌بندی کرد. به خاطر وجود هم‌بستگی‌های قوی و بلند برد بین متغیرهای سامانه، مخصوصاً در نزدیکی گذار فازها، به دست آوردن یک پاسخ تقریبی خوب هم معمولاً از نظر محاسباتی کار دشواری است. مفاهیم و ابزارهای فیزیک آماری سامانه‌های نامنظم به ما کمک کرده‌اند تا الگوریتم‌های کارآمد برای مطالعه این نوع مسئله‌ها طراحی کنیم. در این مقاله، با بعضی از الگوریتم‌ها و سازوکارهای مهم بهینه‌سازی آشنا می‌شویم که از مفاهیم فیزیک کلاسیک و کوانتومی استفاده می‌کنند. تمرکز بیشتر بر روی روش‌ها و یافته‌های سال‌های اخیر در این زمینه خواهد بود. به طور مشخص به اهمیت الگوریتم‌های تقویت شده در بهینه‌سازی و نقش تقویت الگوریتم در کاهش اثر خطا در شبیه‌سازی‌های کوانتومی اشاره می‌کنیم. در نهایت چشم اندازی از مسئله‌ها و راهکارهای پیش‌رو برای بررسی آن‌ها ارائه می‌شود.

واژه‌های کلیدی: فیزیک آماری، بهینه‌سازی، الگوریتم‌های تقویت شده

۱. مقدمه

یادگیری با توجه به قیدهایی که برای قاعده‌مندسازی ساختار شبکه (مختصر بودن) و معنی‌دار بودن خروجی‌های شبکه (مثلاً فیزیکی بودن) در نظر می‌گیریم، فراهم می‌شود. به‌طور کلی مسئله برآورده کردن قیده‌ها را می‌توان به صورت مسئله کمینه کردن یک تابع انرژی مناسب در فضای متغیرهای مسئله نوشت، به طوری که هزینه برآورده نشدن هر قید مقدار تابع انرژی را به اندازه‌ای متناسب با اهمیت آن قید افزایش می‌دهد. حالت‌هایی که این تابع انرژی را کمینه می‌کنند فضای پاسخ‌ها را می‌سازند. مشخصه‌یابی این حالت‌ها از نظر آرایش میکروسکوپی، مقدار انرژی، تعداد، و ساختار فضای پاسخ‌ها، صورت بندی کلی از یک مسئله بهینه‌سازی هستند. در ساده‌ترین شکل، یک الگوریتم بهینه‌سازی مقدار کمینه انرژی و یا یک حالت میکروسکوپی از فضای پاسخ‌ها را برمی‌گرداند [۱].

تعداد متغیرهای لازم برای توصیف حالت میکروسکوپی یک

به‌طور بنیادی، حل هر مسئله را می‌توان به جست‌وجو در فضای نمایش‌ها یا آرایش‌های سازگار با اطلاعات موجود تعبیر کرد. هدف، یافتن حالتی با نمایشی مختصر، معنی‌دار و کارآمد است که علاوه بر برآورده کردن شرایط مسئله، به سادگی قابل محاسبه باشد. برای مطالعه رفتار تعادلی یک سامانه ترمودینامیکی به دنبال تابع توزیع احتمالی از حالت‌های میکروسکوپی می‌گردیم که انرژی آزاد سامانه را کمینه کند و با آن بتوان به راحتی رفتارهای ماکروسکوپی سامانه را توضیح داد. این سامانه می‌تواند الگویی برای یک ماده مغناطیسی در آزمایشگاه باشد یا الگویی از یک پروتئین که در تعادل ترمودینامیکی با محیط اطرافش است. در یک شبکه عصبی مصنوعی هم نمایش بهینه از یک مسئله با کمینه کردن خطای

سامانه (مثلاً جهت اسپین ها یا موقعیت مکانی اتم‌ها) برای سامانه‌هایی که رفتار جالبی دارند معمولاً عدد بزرگی است و از طرفی فضای جستجو هم به‌طور نمایی با این تعداد بزرگ می‌شود. برای مثال، یک پروتئین نوعی حدود چند صد مولکول آمینو اسید را شامل می‌شود که در یک توصیف درشت دانه، مکان هر کدام از آن‌ها در سه بعد حالت سامانه را مشخص می‌کند. از آن طرف، فضای پاسخ‌ها یا محدود است و یا اندازه آن نسبت به اندازه فضای جستجو به‌طور نمایی با تعداد متغیرها کاهش می‌یابد. همین موضوع یافتن یک پاسخ را از نظر محاسباتی برای سامانه‌های بزرگ سخت می‌کند. این که اندازه یک سامانه بزرگ چقدر است را در عمل تعداد و شدت قیدهای مسئله یا تابع انرژی سامانه است که مشخص می‌کند.

پیچیدگی محاسباتی قسمتی از علم کامپیوتر است که سعی می‌کند به‌طور دقیق سختی محاسباتی را از نظر زمان، حافظه، و انرژی لازم برای انجام یک محاسبه کمی کند و مسئله‌های مختلف را از این نظر طبقه‌بندی کند. برای مثال مسئله‌ای از نظر زمان محاسباتی راحت است که در بدترین حالت بتوان آن را با یک تعداد عمل محاسباتی ابتدایی که با تعداد متغیرهای مسئله حداکثر به شکل توانی زیاد می‌شود، حل کرد. در غیر این صورت، به آن یک مسئله سخت از نظر زمان محاسباتی گفته می‌شود. این معیار را پیچیدگی محاسباتی در بدترین حالت می‌گویند. در عمل ما با حالت‌های نوعی از یک مسئله سروکار داریم و نه لزوماً با بدترین نمونه از آن. پیچیدگی محاسباتی نوعی^۱ قرار است سختی محاسباتی نمونه‌هایی از مسئله را بدهد که به‌طور تصادفی و با احتمال مساوی از فضای حالت‌های مسئله انتخاب می‌شوند [۲].

فیزیک آماری جایی است که می‌توان رفتار نوعی تعداد زیادی

متغیر را که یک تابع انرژی به آن‌ها نسبت داده شده، مطالعه کرد. با استفاده از نظریه انحرافات بزرگ^۲ حتی رفتارهای غیرنوعی یک سامانه برهم‌کنشی را نیز بررسی کرد. در این‌جا، برهم‌کنش بین متغیرها توسط قیدهای مسئله که تابعی از متغیرها هستند وارد تابع انرژی سامانه می‌شود. دما نقش ضریب لاگرانژی را بازی می‌کند که با آن می‌توان ناحیه‌های مختلف از فضای حالت‌های میکروسکوپی با انرژی‌های متفاوت را بررسی کرد. افت و خیزهای گرمایی این امکان را فراهم می‌کنند تا در حالت تعادل گرمایی زیر فضای آرایش‌های با انرژی تقریباً یکسان را جستجو کرد. اگر دما به آرامی کاهش یابد، که سامانه همواره فرصت رسیدن به تعادل را داشته باشد، می‌توان دما را به سمت صفر میل داد و از زیر فضای حالت‌های با کمینه انرژی نمونه‌برداری کرد. این‌ها اساس الگوریتم‌های جستجوی کاتوره‌ای با استفاده از افت و خیزهای گرمایی، مانند الگوریتم سرمایه‌سازی شبیه‌سازی شده^۳، هستند [۳]. همین نقش را افت و خیزهای کوانتومی در دمای صفر در الگوریتم‌های جستجوی کوانتومی و سرمایه‌سازی کوانتومی^۴ بازی می‌کنند [۴].

مطالعه سامانه‌های فیزیکی برهم‌کنشی و نامنظم مفاهیمی مانند ناکامی^۵، شکست تقارن نسخه^۶ و روش‌هایی مانند روش نسخه^۷ و روش حفره^۸ برای مطالعه این نوع سامانه‌ها به همراه داشته است. ناکامی در برآورده کردن برهم‌کنش‌های مسئله به‌طور موضعی و شکست تقارن نسخه در سطح‌های مختلف (مرتبه اول، دوم، ...، بینهایت) عامل اصلی پیچیدگی محاسباتی شناخته می‌شوند. به دنبال آن، الگوریتم‌های انتشار باور^۹ معرفی شدند که با رد و بدل کردن پیام‌های موضعی بین متغیرهای سامانه کار می‌کنند [۵ و ۶]. نسخه‌های تقویت‌شده^{۱۰} این الگوریتم‌ها با استفاده از اطلاعاتی که از رفتار آماری خود این الگوریتم‌ها

۱. Typical Complexity

۲. Large Deviations

۳. Simulated Annealing

۴. Quantum Annealing

۵. Frustration

۶. Replica Symmetry Breaking

۷. Replica Method

۸. Cavity Method

۹. Belief Propagation

۱۰. Reinforcement

با کمینه انرژی باشد.

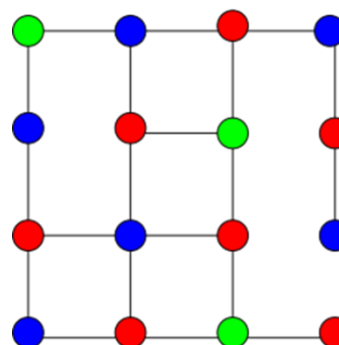
در ادامه این نوشتار، مروری مختصر بر الگوریتم های مهم بهینه سازی خواهیم داشت. با توضیح مفاهیم کلیدی فیزیک آماری سامانه های نامنظم در مسئله های بهینه سازی شروع می کنیم. سپس به طور جداگانه با روش های متداول کلاسیکی و کوانتومی برای بهینه سازی آشنا می شویم. توجه کنید که این مقاله یک مرور جامع از این حوزه نیست، بلکه با توجه به سلیقه و تجربه نویسنده از دیدگاه فیزیک آماری سامانه های نامنظم نوشته شده است.

۲. فیزیک آماری بهینه سازی

برای شروع، مسئله رنگ کردن راس های یک گراف G با تعداد رنگ q را در نظر بگیرید. به ازای هر یال در گراف یک قید داریم، به این معنا که دو راس انتهایی هر یال رنگ های یکسانی نداشته باشند (شکل ۱). یک ریزحالت با رنگ هایی که به راس ها نسبت داده شده مشخص می شود $\vec{\sigma} = \{\sigma_i = 1, \dots, q; i = 1, \dots, N\}$ ، و تابع انرژی تعداد قیدهای برآورده نشده را می دهد:

$$E_G = \sum_{(ij) \in G} \delta_{\sigma_i, \sigma_j}$$

در این رابطه جمع روی تمام یال های گراف است. پاسخ های مسئله حالت هایی هستند که کمترین انرژی را دارند. اگر حالتی با انرژی صفر وجود داشته باشد می گوئیم گراف به اصطلاح رنگ پذیر است. پیدا کردن مقدار کمینه این تابع انرژی در بدترین حالت یک مسئله سخت از نظر زمان محاسباتی است. واضح است که به ازای تعداد راس و رنگ ثابت، اگر چگالی یال ها $\alpha = M / N$ از یک مقدار آستانه α_s بیشتر باشد گراف دیگر رنگ پذیر نخواهد بود. درواقع به ازای $\alpha < \alpha_s$ به احتمال یک در حد ترمودینامیک گراف رنگ پذیر است، درحالی که برای $\alpha > \alpha_s$ این احتمال صفر است. مشاهده می شود که در نزدیکی این گذار زمان محاسبه حتی برای یک گراف نوعی با N راس و M یال هم با اندازه سامانه به شدت زیاد می شود [۱۲].



شکل ۱. یک گراف که راس های آن با سه رنگ قرمز، آبی، و سبز رنگ شده است، به طوری که راس های دو سر هر اتصال رنگ های متفاوت دارند.

به دست آمده روشی نرم برای پیدا کردن حالت های بهینه در مسئله های بهینه سازی است [۷].

خواهیم دید که تقویت الگوریتم های کوانتومی هم در افزایش کارایی آن ها و همچنین در کاهش اثر خطاهای مختلف نقش مهمی می تواند داشته باشد.

از همان ابتدای معرفی شبکه های عصبی مصنوعی روش های فیزیک آماری برای مطالعه کارکرد آن ها استفاده شده است [۸-۱۱]. فازهای ماکروسکوپی الگوهایی مانند هافیلد^۱ و پرسپترون^۲ تک لایه (و حتی چند لایه) به ازای تعداد داده های ورودی و اندازه شبکه عصبی تعیین شد و ظرفیت آن ها در به خطر سپاری الگوهای تصادفی مورد محاسبه قرار گرفت. همچنین، از الگوریتم های بهینه سازی برای پیدا کردن وزن های بهینه برای اتصال های شبکه عصبی و بررسی اثر ساختار فضای پارامترهای شبکه بر یادگیری شبکه های عمیق^۳ استفاده شده است. بررسی عملکرد الگوریتم هایی مانند کاهش گرادیانی^۴ (برای متغیرهای پیوسته) و سرمایه شیب سازی شده کلاسیکی و کوانتومی (برای متغیرهای گسسته) باعث شده که تصویر بهتری از فرآیند یادگیری داشته باشیم. از سوی دیگر، در سال های اخیر از شبکه های عصبی برای یادگیری بهینه سازی استفاده می شود. ورودی شبکه عصبی می تواند یک حالت اولیه پیشنهادی برای پاسخ مسئله بهینه سازی باشد و ساختار شبکه عصبی در فرآیند یادگیری طوری طراحی می شود که خروجی آن یک حالت بهینه

۱. Hopfield

۲. Perceptron

۳. Deep Networks

۴. Gradient Descent

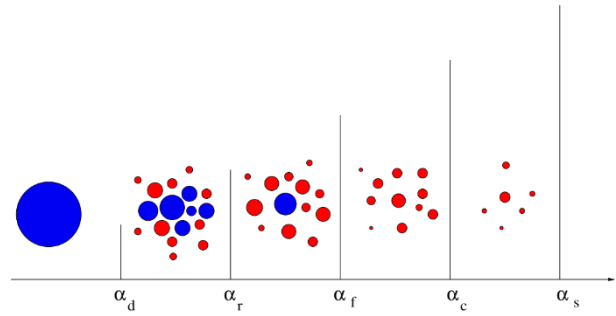
در حد ترمودینامیکی به سمت بینهایت میل می‌کند و کار الگوریتم‌های موضعی را که سعی می‌کنند از فضای حالت‌های تعادلی نمونه برداری کنند، سخت می‌کند. به نظر می‌رسد که این سد را بتوان تا حدی با استفاده از الگوریتم‌های غیرتعادلی و غیرموضعی تا جایی که هزینه‌های محاسباتی قابل قبول باشند، رد کرد [۱۳].

در یک گذار فاز ترمودینامیکی به ازای $\alpha_d \leq \alpha_c \leq \alpha_s$ ، یک تکنیکی در تابع آنتروپی آزاد سامانه ظاهر می‌شود و هم‌بستگی‌های قوی و سرتاسری بین متغیرهای سامانه خواهیم داشت که بیشتر از پیش عملکرد الگوریتم‌های موضعی را کاهش می‌دهند. این تغییر رفتار ماکروسکوپی در حد ترمودینامیکی را یک گذار فاز آرمانی شیشه‌ای می‌نامند. ناکامی روش‌های موضعی به خاطر هم‌بستگی‌های قوی و بلند برد عامل اصلی پیچیدگی‌های محاسباتی در مسئله‌های بهینه‌سازی است. طراحی الگوریتم‌های کارآمدی^۱ که بتوانند نزدیک به گذار رنگ‌پذیری یک پاسخ تقریبی خوب (با انرژی از مرتبه یک) برای مسئله‌های به اندازه کافی بزرگ به دست آورند یک چالش مهم در این زمینه است.

در این جا، ساختار گراف یک بی‌نظمی منجمد شده^۲ در مسئله است، زیرا مسئله رنگ‌پذیری را به ازای یک گراف داده شده بررسی کردیم. برای این که حرفی راجع به رفتار هنگردی از گراف‌های تصادفی با N راس و M یال بزنیم، می‌توانیم با میانگین آنتروپی آزاد کار کنیم که یک کمیت فزون‌ور و در نتیجه خودمیانگین است [۱۴]. در روش نسخه میانگین، کمیت‌هایی مانند آنتروپی آزاد با معرفی تعدادی نسخه مشابه از سامانه و یک فرآیند حدگیری محاسبه می‌شوند:

$$\overline{\Phi}_G = \frac{1}{N} \ln Z_G = \frac{1}{N} \lim_{n \rightarrow \infty} \frac{\overline{Z^n} - 1}{n}$$

این روش تصویری فیزیکی از نقش بی‌نظمی، ناکامی، و ساختار چشم انداز انرژی^۳ در توضیح رفتار نوعی سامانه‌های نامنظم به دست می‌دهد و با معرفی پارامترهای نظم مناسب و مفاهیمی همچون شکست تقارن نسخه و فاز شیشه‌ای توصیفی کارآمد



شکل ۲. تغییر ساختار زیرفضای پاسخ‌ها در فضای آرایش‌های سامانه با افزایش تعداد قیدهای مسئله. در گذار دینامیکی تک خوشه پاسخ‌ها به تعدادی نمایی خوشه کوچکتر می‌شکند و در گذار فاز ترمودینامیکی یک تکنیکی در تابع آنتروپی آزاد ظاهر می‌شود.

در دمای صفر می‌توان یک تابع پارش و آنتروپی آزاد به این سامانه نسبت داد:

$$Z_G = \sum_{\vec{\sigma}} \prod_{(ij) \in G} (1 - \delta_{\sigma_i, \sigma_j}) = e^{N\Phi_G}$$

دقت کنید که تابع پارش در دمای صفر تعداد پاسخ‌های مسئله

$$\Phi_G = \frac{1}{N} \ln Z_G$$

را می‌دهد و به همین خاطر به کمیت Φ_G آنتروپی آزاد گفته می‌شود. حتی قبل از گذار رنگ‌پذیری در این مسئله می‌توان گذارهای ماکروسکوپی دیگری مانند گذار فاز دینامیکی و ترمودینامیکی داشت. بر خلاف گذار فازهای ترمودینامیکی، در گذارهای دینامیکی کمیت‌های تعادلی تکنیکی نشان نمی‌دهند و فقط رفتار دینامیکی سامانه تغییر می‌کند. به‌ویژه، در این جا زمان به تعادل رسیدن سامانه در حد ترمودینامیکی و اگر می‌شود. در یک گذار فاز دینامیکی به ازای $\alpha_d \leq \alpha_s$ فضای پاسخ‌های مسئله شکسته شده و به توزیعی از خوشه‌های با تعداد پاسخ‌های مختلف و دور از هم در فضای آرایش‌های سامانه تبدیل می‌شود (شکل ۲). خوشه‌های پاسخ، به همراه تعدادی نمایی از کمینه‌های موضعی، با فاصله‌ها و دیواره‌های انرژی فزون‌ور از هم جدا شده‌اند. این تغییر ساختار چشم انداز انرژی به یک ساختار پیچیده را شکست تقارن نسخه می‌گویند، زیرا نسخه‌هایی از سامانه که از فضای پاسخ‌ها نمونه‌برداری می‌شوند دیگر از نظر آماری شباهت کمتری با هم دارند. در نتیجه این تغییر ساختار، زمان به تعادل رسیدن سامانه

۱. Efficient Algorithms

۲. Quenched Disorder

۳. Energy Landscape

از رفتار تعادلی و حتی غیرتعادلی این نوع سامانه ها فراهم می کند. این ابزارها و مفاهیم در طیف وسیعی از سامانه های پیچیده که رفتار شبه شیشه ای از خود نشان می دهند مانند مواد آمورف، پروتئین ها، و بازارهای مالی استفاده شده اند.

روش بته^۱ (یا حفره) در فیزیک آماری راهی دیگر برای بررسی این مسئله هاست [۱۴]. در این روش یک تابع وردشی برای آنتروپی آزاد سامانه بر حسب تابع های احتمال حفره^۲ $\mu_{i \rightarrow j}(\sigma_i)$ نوشته می شود که برای یک گراف با ساختار درختی دقیق است:

$$N\Phi_G = \sum_i \ln Z_i - \sum_{(ij) \in G} \ln Z_{ij}$$

$$Z_{ij} = \sum_{\sigma_i, \sigma_j} (1 - \delta_{\sigma_i, \sigma_j}) \mu_{i \rightarrow j}(\sigma_i) \mu_{j \rightarrow i}(\sigma_j)$$

$$Z_i = \sum_{\sigma_i} \prod_{j \in \partial i} [\sum_{\sigma_j} (1 - \delta_{\sigma_i, \sigma_j}) \mu_{j \rightarrow i}(\sigma_j)]$$

تابع توزیع $\mu_{i \rightarrow j}(\sigma_i)$ احتمال آن است که متغیر i ، در غیاب برهم کنش با متغیر j ، در حالت σ_i باشد. در این رابطه ∂_i مجموعه همسایه های راس i است. پیشینه کردن آنتروپی آزاد وردشی نسبت به احتمال های حفره با در نظر گرفتن قید بهنجارش آن ها، معادلات بته را می دهد. این معادلات بازگشتی و تعمیم هایی که شکست تقارن نسخه را هم در نظر می گیرند چارچوبی برای طراحی الگوریتم های انتشار باور فراهم کرده اند که می توان از آن ها برای پیدا کردن پاسخ و مشخصه یابی ساختار فضای پاسخ ها برای یک گراف مشخص استفاده کرد.

هر نقطه ثابت^۳ از معادلات بته را معادل یک حالت گیس خالص^۴ برای سامانه می گیرند، حالتی که در آن هم بستگی بین متغیرها کوتاه برد است. تا زمانی که رفتار نوعی پاسخ های مسئله تنها با یک نقطه ثابت توصیف می شود ما تقارن نسخه داریم. اما معادلات بته می توانند تعدادی نمایی $e^{N\Sigma}$ نقطه ثابت داشته باشند که با هم رفتار آماری سامانه را توضیح می دهند. در این حالت تقارن نسخه شکسته شده است. به کمیت Σ پیچیدگی مسئله گفته می شود. بسته به ساختار سلسه مراتبی که این نقاط

ثابت در فضای حالت های سامانه به خود می گیرند، این تقارن می تواند یک، دو،...، یا بینهایت بار شکسته شده باشد. گذار ترمودینامیکی به فاز شیشه ای در حد ترمودینامیکی با یک تکنیکی در تابع آنتروپی آزاد ظاهر می شود. تعداد نقاط ثابت مربوط به فاز ترمودینامیکی شیشه متناهی است و متغیرهای سامانه در هر پاسخ به شدت همبسته هستند.

مشاهده می کنیم که گذار فازهای ممکن در یک مسئله بهینه سازی اثر زیادی بر روی کارایی الگوریتم ها دارند. اطلاعاتی در مورد مکان و نوع این گذارها کمک می کنند تا تصویر بهتری از موفقیت یا شکست الگوریتم های مختلف در یک مسئله داشته باشیم و حتی بتوانیم پیشاپیش حد بالایی برای عملکرد آن ها برآورد کنیم. از سوی دیگر، با توجه به قضیه نهان^۵ مجانی وجود ندارد می دانیم که قرار نیست یک الگوریتم همه جا خوب کار کند [۱۵]. این اهمیت استفاده از اطلاعات یک مسئله را برای طراحی الگوریتم های کارآمد بیشتر می کند. در ادامه به معرفی تعدادی از الگوریتم های مهم کلاسیکی و کوانتومی در این زمینه می پردازیم. هر یک از این روش ها محدودیت های خود را دارد ولی کار کردن با آن ها در بیشتر مسئله ها راحت است که این موجب فراگیرتر شدن آن ها شده است.

۳. الگوریتم های بهینه سازی کلاسیکی

۳.۱. الگوریتم های جستجوی کاتوره ای

ساده ترین روش جستجو، انتخاب تصادفی یک حالت از فضای آرایش های سامانه و محاسبه انرژی آن است. اگر این حالت پاسخ مسئله باشد، فرایند پایان می یابد؛ در غیر این صورت، حالت تصادفی دیگری انتخاب می شود. ولی تعداد پاسخ ها معمولاً به طور نمایی کوچکتر از تعداد کل حالت های سامانه است، در نتیجه متوسط زمان لازم برای پیدا کردن یک پاسخ به طور نمایی با تعداد متغیرها زیاد می شود. با این حال، می توان

۱. Bethe

۲. Cavity Probabilities

۳. Fixed Point

۴. Pure Gibbs State

۵. No-Free-Lunch Theorems

وزن‌های اتصال‌های شبکه $\mathbf{w}$ است. تغییرات وزن‌ها با متغیر زمان پیوسته در فرآیند یادگیری را می‌توان به شکل زیر نوشت:

$$\frac{\partial}{\partial t} \mathbf{w} = -\eta \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}) + \epsilon(t)$$

اندازه ضرب یادگیری $\eta(t)$ و شدت نوفه $\epsilon(t)$ می‌توانند تابعی کاهشی از زمان باشند تا با گذشت زمان سامانه به یک حالت ثابت برسد. همچنین نشان داده شده است که در نظر گرفتن آنتروپی پاسخ‌ها در تابع خطا می‌تواند به پیدا کردن پاسخ‌هایی که در کمینه‌های موضعی پهن^۳ قرار دارند، کمک کند [۱۷ و ۱۸]. این ویژگی پاسخ‌ها در شبکه‌های عصبی عمیق توضیحی برای قابلیت تعمیم^۴ بالای شبکه‌های عمیق با وجود پارمترهای زیاد است.

شباهت معادله بالا با معادله لانژون^۵ نیز امکان تحلیل رفتار الگوریتم گرادیان کاهشی را با استفاده از ابزار و مفاهیم فیزیک آماری فراهم می‌کند [۱۹ و ۲۰]. از این نظر تابع خطای یادگیری نقش پتانسیل خارجی را بازی می‌کند و افت و خیزهای گرمایی را نوفه اضافه شده به معادله نمایندگی می‌کند. انتظار می‌رود که اضافه کردن جمله اینرسی (مشتق زمانی دوم) به معادله بالا تغییرات زمانی متغیرها را هموارتر کند.

۳.۲. الگوریتم‌های انتشار باور

معادلات بته برای احتمال‌های حفره را در مسئله رنگ‌آمیزی گراف در نظر بگیرید که با کمینه کردن آنتروپی آزاد وردشی به دست می‌آیند [۲۱ و ۲۱]:

$$\mu_{i \rightarrow j}(\sigma_i) = \frac{1}{Z_{i \rightarrow j}} \prod_{k \in \partial i \setminus j} [\sum_{\sigma_k} (1 - \delta_{\sigma_i, \sigma_k}) \mu_{k \rightarrow i}(\sigma_k)]$$

که در آن $Z_{i \rightarrow j}$ ثابت بهنجارش است. فرض اساسی در این معادلات این است که متغیرهای همسایه در غیاب برهم‌کنش با متغیر i تقریباً مستقل از یکدیگر هستند (شکل ۳).

کمی هدفمندتر از این رفتار کرد. بگذارید از یک حالت اولیه در فضای آرایش‌ها شروع کنیم و آن را کمی تغییر دهیم $(\sigma' = \sigma + \delta\sigma)$. اگر انرژی سامانه کاهش پیدا کرد $(\Delta E = E' - E < 0)$ آرایش جدید پذیرفته می‌شود $(\sigma \rightarrow \sigma')$ ، در غیر این صورت تغییر دیگری را امتحان می‌کنیم. این کار را تا جایی ادامه می‌دهیم که دیگر حالتی با انرژی کمتر پیدا نکنیم. با این روش حتماً یک کمینه موضعی^۱ از تابع انرژی را پیدا خواهیم کرد.

برای این که در این کمینه‌های موضعی گرفتار نشویم می‌توان کمی نوفه به سامانه اضافه کرد و با احتمال $e^{-\beta \Delta E}$ تغییراتی را که موجب افزایش انرژی می‌شوند نیز پذیرفت. متغیر β نقش معکوس دما را در یک سامانه ترمودینامیکی بازی می‌کند و امکان عبور از سدهای انرژی بین کمینه‌های موضعی را فراهم می‌کند. الگوریتم شبیه‌سازی سرمایش از این شباهت استفاده می‌کند تا با کاهش دما، سامانه را به سمت حالت‌های با انرژی کمینه برساند [۳]. با شروع از یک دمای به اندازه کافی بالا و کاهش آهسته آن به طوری که در هر قدم سامانه تقریباً در حالت تعادل باشد می‌توان از سدهای انرژی با اندازه‌های مختلف عبور کرد و سامانه را به حالت‌های تعادلی دمای صفر یعنی حالت‌های با کمترین انرژی رساند. متأسفانه کار به این سادگی نیست و با کاهش دما گذار فازهایی اتفاق می‌افتند که زمان به تعادل رسیدن را به طور نمایی با تعداد متغیرها زیاد می‌کنند. این مشکل اهمیت الگوریتم‌های غیرتعادلی را برای دور زدن یا کاهش اثر گذار فازها و نمونه برداری از حالت‌های نادر بیشتر می‌کند [۱۶].

در خیلی از مسئله‌ها مانند یادگیری در شبکه‌های عصبی مصنوعی و حل معادلات ماتریسی، با متغیرهای پیوسته سروکار داریم. در این شرایط اگر تابع انرژی مشتق پذیر باشد می‌توان از الگوریتم کاهش گرادیانی کاتوره‌ای^۲ برای پیدا کردن پاسخ‌های تقریبی استفاده کرد. برای مثال تابع خطای یادگیری $\mathcal{L}(\mathbf{w})$ در یک شبکه عصبی را در نظر بگیرید که تابعی از

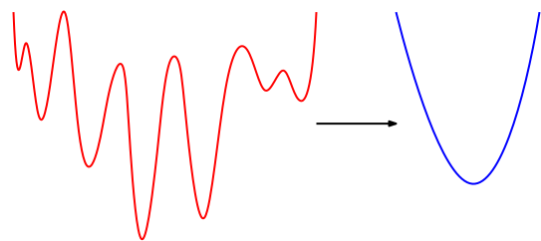
۱. Local Minimum

۲. Stochastic Gradient Descent

۳. Flat Local Minima

۴. Generalization

۵. Langevin



شکل ۴. تقویت الگوریتم بهینه سازی به تدریج چشم انداز انرژی را ساده تر می کند و در نتیجه کار پیدا کردن حالت بهینه راحت تر می شود.

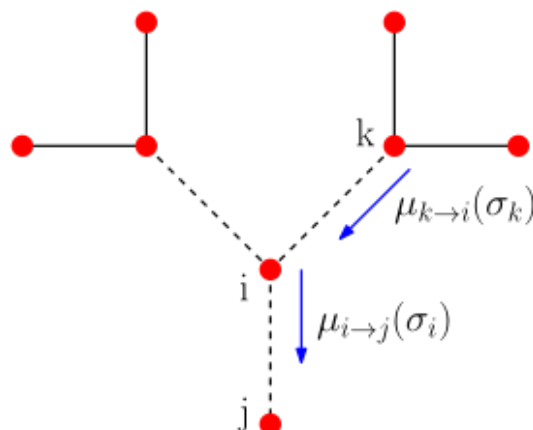
$$\mu_{i \rightarrow j}^{t+1}(\sigma_i) = \frac{e^{h_i^t(\sigma_i)}}{Z_{i \rightarrow j}} \prod_{k \in \partial i \setminus j} [\sum_{\sigma_k} (1 - \delta_{\sigma_i, \sigma_k}) \mu_{k \rightarrow i}^t(\sigma_k)]$$

$$p_i^{t+1}(\sigma_i) = \frac{e^{h_i^t(\sigma_i)}}{Z_i} \prod_{j \in \partial i} [\sum_{\sigma_j} (1 - \delta_{\sigma_i, \sigma_j}) \mu_{j \rightarrow i}^t(\sigma_j)]$$

$$h_i^{t+1}(\sigma_i) = r \ln p_i^{t+1}(\sigma_i)$$

پارامتر تقویت کننده r را می توان در ابتدا صفر در نظر گرفت و سپس به تدریج افزایش داد تا همزمان با دقیق تر شدن احتمال های موضعی، سامانه بیش از پیش به سمت پاسخ های مسئله هدایت شود. دقت شود که میدان تقویت کننده، چشم انداز انرژی سامانه را به طور پیوسته تغییر می دهد، به گونه ای که در نهایت یک تابع توزیع احتمال بدیهی خواهیم داشت که تنها برای یکی از پاسخ های مسئله مقدار قابل ملاحظه ای دارد (شکل ۴).

الگوریتم بالا را می توان به حالتی تعمیم داد که تقارن نسخه شکسته شده باشد. در این شرایط انتظار داریم که معادلات به بیشتر از یک پاسخ داشته باشند. هر پاسخ معادل یک حالت گیس خالص از تابع توزیع احتمال سامانه است. در یک حالت گیس خالص هم بستگی ها به صورت نمایی با فاصله کم می شوند و در نتیجه احتمال های حفره تخمین خوبی از حالت آماری سامانه ارائه می کنند. معادلات به در این جا در یک سطح بالاتر و برای توزیع احتمال های حفره نوشته می شوند، با این فرض که تقارن نسخه یک بار شکسته شده است. وقتی از روش تکرار برای حل این معادلات استفاده می شود به آن الگوریتم انتشار جمع بندی باورها^۳ می گویند. همانند قبل می توان در این جا از تقویت الگوریتم برای پیدا کردن پاسخ های تقریبی



شکل ۳. متغیرهای مسئله با انتشار باورهای خود از حالتی که با توجه به قیدهای مسئله و حالت دیگر متغیرها می توانند اختیار کنند، به سمت فضای پاسخ ها میل می کنند.

برای شبکه های درهم کنشی که به طور موضعی ساختار درختی دارند، فرض استقلال متغیرهای همسایه تقریب بسیار خوبی است. معادلات بازگشتی بالا را می توان با روش تکرار و با شروع از احتمال های حفره تصادفی حل کرد. این روش با انتقال پیام های حفره $\mu_{i \rightarrow j}(\sigma_i)$ بین متغیرهای همسایه در یک شبکه درختی به تنها پاسخ معادلات، که پاسخی دقیق هم است، همگرا می شود. اگر شبکه درهم کنشی درختی نباشد معادلات به (انتشار باور) می توانند چند پاسخ داشته باشند و لزومی نیز ندارد که روش تکرار همگرا شود. در صورتی که یک پاسخ به دست آید، تخمینی از احتمال های موضعی^۱ را می شود با استفاده از احتمال های حفره حساب کرد:

$$p_i(\sigma_i) = \frac{1}{Z_i} \prod_{j \in \partial i} [\sum_{\sigma_j} (1 - \delta_{\sigma_i, \sigma_j}) \mu_{j \rightarrow i}(\sigma_j)]$$

اگر هدف تنها یافتن یک پاسخ تقریبی مناسب باشد، می توان از اطلاعات موضعی بالا برای تقویت الگوریتم انتشار باور استفاده کرد. برای این کار یک میدان خارجی به شکل $h_i(\sigma_i) = r \ln p_i(\sigma_i)$ تعریف می شود تا متناسب با احتمال به دست آمده در هر قدم t از روش تکرار، بر جهت گیری متغیرها در قدم بعدی $t+1$ اثر بگذارد. به این ترتیب معادلات تقویت شده به ^۲را داریم:

۱. Local Probabilities

۲. Reinforced Bethe Equations

۳. Survey Propagation

بالا) از آن تولید می کنند. این فرآیند را با تعمیم متغیرهای مسئله σ به متغیرهای پیوسته x می توان، برای مثال، با یک تابع درونیابی خطی نمایش داد:

$$x(t) = \alpha(t)\sigma + \beta(t)z$$

ضرایب $\alpha(t), \beta(t)$ به گونه ای تعیین می شوند که در زمان $t=0$ و $t=1$ به ترتیب پاسخ داده شده $\sigma = x(0)$ و یک آرایش کاملاً تصادفی $x(1) = z$ تولید شود. در فرایند معکوس، زمان به عقب برمی گردد و با شروع از یک حالت تصادفی در $t=1$ یک پاسخ تقریبی در $t=0$ تولید می شود. برای این کار باید در فرآیند رو به عقب با بی اثر کردن نوفه در هر بازه زمانی حالت هایی را به دست آورد که در فرآیند رو به جلو در آن بازه زمانی اتفاق افتاده اند. در ضمن هر قدم از فرآیند وارون را می توان با یادگیری یک شبکه عصبی الگوسازی کرد [۲۶ و ۲۷].

هنوز مشخص نیست که شبکه های عصبی تا چه اندازه می توانند عملکرد بهتری از الگوریتم های شناخته شده بهینه سازی داشته باشند. همچنین، آیا می توان با نگاه کردن به فرآیند یادگیری و دینامیک لایه های پنهان شبکه عصبی یک الگوریتم بهینه سازی معنی دار طراحی کرد؟

۴. الگوریتم های بهینه سازی کوانتومی

در این بخش به طور مختصر به معرفی تعدادی از الگوریتم های بهینه ساز کوانتومی می پردازیم. برای مرور جامع تر این موضوع، خواننده به مرجع [۲۸] ارجاع داده می شود.

۴.۱. جستجوی کوانتومی

فرض کنید که به دنبال یک آرایش مشخص از بین $D = 2^N$ حالت هستیم که با رشته هایی به طول N از ± 1 کدبندی شده اند. از نظر کلاسیکی زمانی از مرتبه D نیاز داریم تا حالت هدف را در این فضا پیدا کنیم. از طرفی این حالت ها را می شود پایه های محاسباتی فضای هیلبرت یک سامانه کوانتومی

مسئله بهره برد. این روش نسبت به الگوریتم انتشار باور اطلاعات بیشتری از فضای پاسخ های مسئله جمع آوری می کند و در نتیجه انتظار می رود که با سخت تر شدن مسئله در نزدیکی گذار فازها عملکرد بهتری داشته باشد [۶].

الگوریتم های انتشار جمع بندی باور را می توان در مسئله های با متغیرهای پیوسته نیز به کار برد. در این چارچوب آن ها را با نام الگوریتم های انتشار باور تقریبی^۱ می شناسند [۲۲]. اهمیت این روش ها در کارایی آن ها برای شبکه های برهم کنشی^۲ تنگ است، جایی که زمان محاسبه به صورت خطی با تعداد متغیرها زیاد می شود. حتی زمانی که شبکه برهم کنشی متغیرها تنگ نیست، نمایش های کارآمد و تقریبی از الگوریتم های انتشار باور می توانند درک عمیق تری از مسئله های بهینه سازی فراهم کنند [۲۳].

۳.۳. یادگیری بهینه سازی

الگوریتم هایی که در بالا گفته شد بر اساس سازوکاری طراحی شده اند که قابل تعبیر و معنی دار هستند. فرض کنید که یک شبکه عصبی مصنوعی قرار است پاسخ مسئله بهینه سازی را پیدا کند. ورودی شبکه عصبی می تواند یک آرایش تصادفی از متغیرهای مسئله و خروجی یک پاسخ تقریبی برای مسئله باشد. ساختار داخلی لایه های شبکه عصبی را شبکه برهم کنشی متغیرها تعیین می کند. در نتیجه این شبکه ها با نام شبکه های عصبی گرافی^۳ شناخته می شوند [۲۴ و ۲۵]. تابع خروجی هر نورون را می توان متناسب با تابع انرژی مسئله مشخص کرد. در فرآیند یادگیری وزن های اتصال های شبکه را به گونه ای تغییر می دهیم تا انرژی حالت خروجی را کمینه کنند. به این ترتیب الگوریتم های انتشار باور را می توان نمایش خاصی از شبکه های عصبی گرافی دانست که در آن وزن های شبکه توسط معادلات بته تعیین می شوند.

بعضی از الگوهای مولد^۴ با یک پاسخ داده شده از مسئله شروع می کنند (معادل دمای پایین) و در یک فرآیند رو به جلو با اضافه کردن نوفه به پاسخ، یک آرایش کاملاً تصادفی (معادل دمای

۱. Approximate Message Passing Algorithms

۲. Sparse

۳. Graph Neural Networks

۴. Generative Models

$\{\beta_i, \gamma_i\}$ به گونه ای تعیین شوند که حالت نهایی تا حد امکان به حالت هدف نزدیک باشد. تحول های یکانی و مقدارهای چشم داشتی حالت نهایی را می توان با استفاده از یک سامانه کوانتومی انجام داد و مسئله پیدا کردن پارامترهای بهینه را به یک الگوریتم کلاسیکی سپرد. واضح است که سختی مسئله به صورت نمایی با تعداد لایه ها L رشد می کند. در ضمن احتمال موفقیت این روش خیلی بستگی به هامیلتونی های H_x, H_p مورد استفاده در عملگرهای یکانی دارد. انتخاب مناسب این هامیلتونی ها می تواند سختی محاسباتی پیدا کردن پارامترهای بهینه را کاهش دهد [۳۲-۳۴]. در ادامه، بیشتر به الگوریتم سرمایه گذاری کوانتومی^۴ خواهیم پرداخت که چارچوب نظری برای روش هایی شبیه به الگوریتم های بالا فراهم می آورد.

۲.۴. سرمایه گذاری کوانتومی

از ترمودینامیک کلاسیک می دانیم که تغییرات شبه ایستا^۵ و بدون اتلاف، آنتروپی کل را تغییر نمی دهند. در یک سامانه کوانتومی می توان از حالت پایه یک هامیلتونی بدیهی H_i شروع کرد و خیلی آهسته هامیلتونی را در دمای صفر به H_f تغییر داد، به طوری که در تمام فرآیند سامانه در حالت پایه هامیلتونی لحظه ای باشد [۴، ۳۵ و ۳۶]. قضیه بی دررو^۶ بیان می کند که زمان این محاسبه متناسب با معکوس مربع کمترین گاف انرژی Δ_{min} در طی این فرآیند است (شکل ۵). مسئله این است که با تغییر هامیلتونی یک گذار فاز کوانتومی گسسته می تواند اتفاق بیافتد که در آن گاف انرژی به صورت نمایی با اندازه سامانه کاهش می یابد. اگر محل وقوع این گذار فاز از پیش معلوم باشد، می توان نرخ تغییر هامیلتونی را در نزدیکی آن کاهش داد و در نواحی دورتر سریع تر هامیلتونی را تغییر داد تا از این طریق زمان محاسبه کاهش یابد. با این حال، در مسئله های واقعی معمولاً محل گذار فاز از پیش مشخص نیست و تعیین آن نیز خود مسئله ای دشوار است.

از N کیوبیت در نظر گرفت. الگوریتم کوانتومی گروور^۱ این زمان محاسبه را به $\sqrt{D}$ کاهش می دهد [۲۹-۳۱]. بگذارید حالت هدف را با $|\psi_f\rangle$ نشان دهیم. حالت اولیه ترکیب خطی همه حالت ها را می توان به صورت $|\psi_i\rangle = \cos(\theta/2)|\psi_f^\perp\rangle + \sin(\theta/2)|\psi_f\rangle$ نوشت، که در آن $\sin(\theta/2) = 1/\sqrt{D}$ و $|\psi_f^\perp\rangle$ عمود بر حالت هدف است. در هر قدم از الگوریتم گروور، یک عمل وارون سازی نسبت به حالت اولیه و حالت هدف انجام می شود که با عملگر زیر نمایش داده می شود:

$$G = -(\mathbb{I} - |\psi_i\rangle\langle\psi_i|)(\mathbb{I} - |\psi_f\rangle\langle\psi_f|)$$

مشاهده می شود که عملگر بالا معادل یک دوران به اندازه زاویه θ در زیرفضای دو بعدی $\{|\psi_f\rangle, |\psi_f^\perp\rangle\}$ است. بنابراین، پس از L مرحله، خواهیم داشت:

$$G^L |\psi_i\rangle = \cos([2L+1]\theta/2) |\psi_f^\perp\rangle + \sin([2L+1]\theta/2) |\psi_f\rangle$$

حداقل تعداد عمل لازم (زمان محاسبه) برای رسیدن به حالت هدف از شرط $|\sin([2L+1]\theta/2)|^2 \approx 1$ به دست می آید که رابطه مقیاسی معروف $L_{min} \propto \sqrt{D}$ را می دهد.

الگوریتم جستجوی کوانتومی^۲ را می توان به حالتی تعمیم داد که هدف یافتن حالت پایه هامیلتونی H_p از یک سامانه N کیوبیتی باشد. در الگوریتم بهینه سازی تقریبی کوانتومی^۳ با یک حالت

$$|\psi_0\rangle = \prod_{i=1}^N \left(\frac{|+\rangle_i + |-\rangle_i}{\sqrt{2}} \right)$$

می شود که حالت پایه هامیلتونی $H_x = -\sum_{i=1}^N \sigma_i^x$ است. در

این جا، $|\pm\rangle_i$ ویژه حالت های عملگر پائولی σ_i^z هستند. در هر قدم دو عملگر یکانی $U_x = e^{-i\beta H_x}, U_p = e^{-i\gamma H_p}$ با پارامترهای β, γ بر حالت سامانه عمل می کنند. این فرایند را برای L بار تکرار می کنیم، $|\psi_{l+1}\rangle = U_p(\gamma_l)U_x(\beta_l)|\psi_l\rangle$ ، تا حالت نهایی $|\psi_L\rangle$ را به دست آوریم. قرار است پارامترهای

۱. Grover

۲. Quantum Search Algorithm

۳. Quantum Approximate Optimization Algorithm

۴. Quantum Annealing

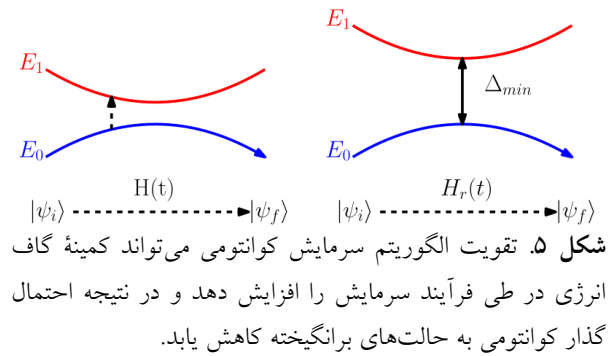
۵. Quasi Static

۶. Adiabatic Theorem

حالت‌های کمینه را آسان‌تر کرد. در نتیجه انتظار داریم که تقویت مناسب الگوریتم سرمایه‌ش کوانتومی هم با تخصیص انرژی‌های کمتر به حالت‌های محتمل‌تر گاف انرژی را افزایش دهد و از این طریق زمان محاسبه را کاهش دهد (شکل ۵). به‌طور دقیق‌تر اگر حالت کوانتومی سامانه $|\psi(t)\rangle$ را داشته باشیم، آنگاه از هامیلتونی تقویت شده زیر برای شبیه‌سازی سرمایه‌ش کوانتومی استفاده می‌کنیم:

$$H_r(t) = A(t)H_i + B(t)H_f - r(t) \sum_{\sigma} \ln |\psi_t(\sigma)|^2 |\sigma\rangle\langle\sigma|$$

در این رابطه $r(t)$ ضریب تقویت کننده است و $\psi_t(\sigma) = \langle\sigma|\psi(t)\rangle$ تصویر تابع موج لحظه‌ای است که در پایه محاسباتی $|\sigma\rangle$ نوشته شده است. ضرایب $A(t), B(t)$ همانند قبل به ترتیب تابعی کاهشی و افزایشی با زمان هستند. نشان داده شده است که تقویت کردن الگوریتم سرمایه‌ش کوانتومی در مسئله‌های بهینه‌سازی مختلف، مانند پیدا کردن حالت پایه الگوی آیزینگ با برهم‌کنش‌های تصادفی، کمینه گاف انرژی را افزایش می‌دهد [۳۹]. حتی در یک الگوی میدان متوسط که گذار فاز کوانتومی گسسته دارد اضافه کردن جمله تقویت کننده می‌تواند گذار فاز را به پیوسته تغییر دهد و در نتیجه رفتار گاف انرژی با اندازه سامانه را از نمایی به توانی تبدیل کند [۴۰]. به‌علاوه، در تعدادی از مسئله‌های بهینه‌سازی مشاهده می‌کنیم که احتمال پیدا کردن پاسخ به اندازه قابل ملاحظه‌ای با تقویت الگوریتم زیاد می‌شود. اما مشکل تقویت الگوریتم‌های کوانتومی این است که شما نیاز دارید اطلاعاتی از حالت کوانتومی سامانه در هر لحظه داشته باشید، در حالی که هر نوع اندازه‌گیری حالت کوانتومی را تا حدی مختل می‌کند. در اصل می‌توان با اندازه‌گیری ضعیف دسته جمعی^۱ از تعداد زیادی نسخه مشابه از سامانه، بدون آن‌که حالت سامانه را زیاد تغییر دهیم، اطلاعاتی از آن به‌دست آورد [۴۱]. ولی این کار نیز هزینه محاسباتی بالایی دارد و جمع شدن خطاها نیز در طی فرآیند می‌تواند کارایی نهایی الگوریتم را کاهش دهد. باید دید که چطور می‌شود راه کارهایی برای غلبه بر این مشکلات



روش‌های متعددی برای کاهش اثر گذارهای کوانتومی پیشنهاد شده است که معمولاً با اضافه کردن جملاتی به هامیلتونی سعی می‌کنند جلوی برانگیخته شدن سامانه را بگیرند و یا از آن‌ها برای رسیدن به حالت هدف استفاده کنند [۳۷ و ۳۸]. مشکل اصلی این روش‌ها آن است که شبیه‌سازی این هامیلتونی‌ها به خاطر جملات غیرموضعی سخت است و تضمینی هم در روش‌های تقریبی برای کار کردن آن‌ها در حالت کلی برای سامانه‌های به اندازه کافی بزرگ وجود ندارد.

فرض کنید قرار است حالت با کمترین انرژی $E(\sigma)$ از یک سامانه کلاسیکی با N اسپین دو حالت $\sigma_i = \pm 1$ را پیدا کنیم. پاسخ‌های این مسئله حالت پایه هامیلتونی $H_f = E(\sigma^z)$ هستند که در آن اسپین‌های کلاسیکی با مولفه z ماتریس‌های پائولی جایگزین شده‌اند. اکنون هامیلتونی وابسته به زمان زیر را در نظر بگیرید:

$$H(t) = A(t)H_i + B(t)H_f$$

به‌گونه‌ای که در زمان صفر $H(0) = H_i$ و برای $t = T$ در پایان فرآیند $H(T) = H_f$ است. هامیلتونی اولیه را می‌توان به صورت $H_i = -\sum_i \sigma_i^x$ با حالت پایه

$$|\psi_i\rangle = \prod_i \left(\frac{|+\rangle_i + |-\rangle_i}{\sqrt{2}} \right)$$

حالت‌های $|\pm\rangle_i = \frac{1}{\sqrt{2}} (|+\rangle_i \pm |-\rangle_i)$ به عنوان پایه‌های محاسباتی هستند. واضح است که در فضای تمامی هامیلتونی‌های ممکن برای سامانه، بینهایت مسیر برای رفتن از H_i به H_f وجود دارد. مسیر بهینه باید بیشترین کمینه گاف انرژی را داشته باشد. از الگوریتم‌های کلاسیکی می‌دانیم که با تقویت حالت‌های با احتمال بیشتر می‌توانیم چشم‌انداز انرژی را تغییر دهیم و یافتن

می شود. در الگوریتم گرور پارامتر τ_l به شکل زیر تعریف می شود:

$$\tau_l = \frac{1}{2} \left[1 - \sqrt{\frac{P_0}{1-P_0}} \tan \left[\left(1 - \frac{2l}{L-1} \right) \theta \right] \right]$$

$$\theta = \arctan \left[\sqrt{\frac{1-P_0}{P_0}} \right]$$

P_0 احتمال اولیه برای پیدا کردن پاسخ است. در این طراحی از این نکته استفاده شده است که گذار فاز کوانتومی در $l^* = (L-1)/2$ اتفاق می افتد تا در نزدیکی آن تغییرات زمانی هامیلتونی خیلی آهسته باشد. استفاده از همین نکته زمان محاسبه یا تعداد لایه ها را از D در فضای هیلبرت $D = 2^N$ بعدی به $\sqrt{D}$ کاهش می دهد.

زمان محاسبه الگوریتم گرور کمترین زمانی است که می توان از یک تحول زمانی خطی (معادله شرودینگر) انتظار داشت. اما یک الگوریتم جستجوی تقویت شده که معادل یک تحول غیرخطی است می تواند زمان محاسبه را به طور نمایی کاهش دهد [۴۴]. در اینجا، زمان محاسبه تعداد لایه های تحول زمانی است که لازم داریم تا احتمال پیدا کردن حالت هدف از یک حد آستانه بیشتر شود.

وقتی زمان محاسبه به طور نمایی از مقدار داده شده در الگوریتم گرور کمتر باشد انتظار داریم که حساسیت الگوریتم به خطا هم به همین نسبت کمتر شود. نشان داده شده است که مقدار آستانه تحمل خطا در الگوریتم گرور با تعداد متغیرها به صورت نمایی کم می شود که با توجه به زمان محاسبه آن انتظار نیز داشتیم [۴۵ و ۴۶]. در نتیجه برای یک زمان محاسبه توانی، آستانه تحمل خطا هم به شکل معکوس توانی خواهد بود.

مشاهده شده است که احتمال موفقیت الگوریتم سرمایه گذاری کوانتومی به ازای خطاهای همدرس و ناهمدوس با تقویت الگوریتم زیاد می شود. فرض کنید که خطا به شکل همدرس در هامیلتونی وارد شده تا حالت سامانه را در یک تحول یکانی تقویت شده منحرف کند. به طور دقیق تر، هامیلتونی تقویت شده در لایه l به شکل زیر است:

$$H_l = A_l H_i + B_l H_f - r \rho_l + \epsilon V_l$$

یافت و در عین حال از مزیت های الگوریتم های تقویت شده نیز بهره برد.

۳.۴. بهینه سازی در حضور خطا

در عمل، وقتی یک الگوریتم کوانتومی روی یک سامانه فیزیکی اجرا می شود با انواع خطاهایی سر و کار داریم که رفتار آرمانی سامانه را بهم می ریزند. جدای از خطاهای آماده سازی و اندازه گیری حالت کوانتومی، در مدت زمان اجرای الگوریتم، خطاهای نظام مندی در هامیلتونی داریم که تحول زمانی سامانه را منحرف می کنند. علاوه بر این، برهم کنش های ناخواسته با محیط را اضافه کنید که به شکل ناهمدوس^۱ بر دینامیک سامانه اثر می گذارند. این خطاها به شدت کارایی الگوریتم های کوانتومی را کاهش می دهند. الگوریتم های تصحیح خطا با استفاده از متغیرهای کمکی در کنار درجات آزادی اصلی مسئله و یا بهره گیری از ویژگی های مقاوم در برابر خطا در سامانه های توپولوژیکی، قادر به کاهش این خطاها هستند. ولی استفاده از آنها مستلزم استفاده از سامانه های بزرگ کوانتومی (چند میلیون کیوبیت) و همچنین آماده سازی حالت های ویژه کوانتومی دارد. از این نظر به کارگرفتن روش های کم هزینه تر برای کاهش اثر خطاها در سال های اخیر اهمیت یافته است [۴۲ و ۴۳].

یکی از الگوریتم های مهمی که حساسیت بالایی به خطا دارد، الگوریتم جستجوی کوانتومی گرور است. این الگوریتم را می توان به شکل یک الگوریتم سرمایه گذاری کوانتومی هم نمایش داد. دوباره سامانه ای شامل N کیوبیت را در نظر بگیرید. در

زمان صفر از حالت بدیهی $|\psi_i\rangle = \frac{1}{\sqrt{2^N}} \sum_{\sigma} |\sigma\rangle$ که حالت

پایه $H_i = \mathbb{I} - |\psi_i\rangle\langle\psi_i|$ است، آغاز می کنیم. حالت کوانتومی هدف $|\psi_f\rangle$ نیز حالت پایه

$H_f = \mathbb{I} - |\psi_f\rangle\langle\psi_f|$ است. دینامیک سامانه با تعداد L

لایه تحول یکانی $|\psi_{l+1}\rangle = e^{-iH_l} |\psi_l\rangle$ با هامیلتونی $H_l = (1 - \tau_l)H_i + \tau_l H_f$ در لایه $l = 0, \dots, L-1$ داده

در این رابطه $\rho_l = |\psi_l\rangle\langle\psi_l|$ حالت کوانتومی سامانه و r ضریب تقویت است. جمله خطا V_l شامل عملگرهای پائولی است که تنها بر روی یک کیوبیت عمل می‌کنند. ضرایب A_l, B_l به گونه‌ای تعیین می‌شوند تا احتمال به دست آوردن پاسخ در هر لایه را بیشینه کنند. تقویت الگوریتم جستجوی کوانتومی به طور قابل ملاحظه‌ای کارایی آن را در حضور خطا افزایش می‌دهد [۴۴]. به نظر می‌رسد که جمله تقویت کننده با تصویر کردن حالت سامانه بر زیر فضای حالت‌های بدون خطا، اثر انحراف دینامیک سامانه را کاهش می‌دهد. با این حال، این که تا چه اندازه می‌توان این نتایج را با هزینه کمتر برای سایر مسئله‌های بهینه‌سازی مشاهده کرد، همچنان موضوع پژوهش‌های آینده است.

۵. جمع‌بندی و چشم انداز

به طور خلاصه مشاهده کردیم که چگونه با استفاده از مفاهیم و ابزارهای فیزیک آماری کلاسیک و کوانتومی، الگوریتم‌های بهینه‌سازی طراحی کرد. این روش‌ها نه تنها کمک می‌کنند تا در بعضی از مسئله‌ها پاسخ‌های تقریبی مناسبی به دست آوریم، بلکه تصویری روشن از ساختار پاسخ‌ها در فضای آرایش‌های سامانه فراهم می‌کنند. این تصویر توضیح می‌دهد که چرا و در چه شرایطی یک الگوریتم در پیدا کردن پاسخ با مشکل مواجه می‌شود و در نتیجه کمک می‌کند تا روش‌هایی برای حل مسئله پیشنهاد کنیم. به ویژه، مشاهده شد که تقویت الگوریتم‌های کلاسیکی و کوانتومی احتمال پیدا کردن پاسخ را افزایش می‌دهد و به کاهش اثر خطاها در الگوریتم جستجوی کوانتومی کمک می‌کند.

در پایان به تعدادی از مسئله‌های مهم پیش رو از دید نویسنده اشاره می‌شود:

مراجع

- به دست آوردن پاسخ‌های تقریبی خوب هنوز هم در خیلی از مسئله‌های بهینه‌سازی مانند رنگ‌آمیزی گراف از نظر محاسباتی دشوار است، به ویژه در نزدیکی گذار فاز شیشه که هم‌بستگی‌های بلند برد ظاهر می‌شوند. تغییر هوشمندانه چشم‌انداز انرژی از طریق روش‌های مناسب تقویت الگوریتم و یا در نظر گرفتن آنتروپی پاسخ‌ها می‌تواند در این زمینه سودمند باشد. می‌توان به روش‌هایی کارآمد برای استفاده از اطلاعات سرتاسری در مورد وضعیت متغیرها اندیشید تا از این طریق تا حدی ناکامی‌های موضعی را برطرف کرد.
- بهینه‌سازی با متغیرهای پیوسته چالش‌های خودش را دارد. پیدا کردن نمایش‌ها و الگوریتم‌های کارآمد برای چنین مسئله‌هایی مانند مسئله چینش با بیشینه چگالی کره‌های سخت در یک فضای اقلیدسی d بعدی، علاوه بر اهمیت نظری در کاربردهایی مانند کدگذاری و انتقال اطلاعات نیز نقش مهمی دارد. الگوریتم‌های انتشار باور تقریبی سعی می‌کنند این مشکل را برطرف کنند.
- کارهای زیادی با الگوریتم سرمایه‌گذاری کوانتومی انجام شده است. با وجود این هنوز نمی‌شود با اطمینان گفت که عملکرد این الگوریتم در مسئله‌های مهم بهینه‌سازی بهتر از الگوریتم‌های کلاسیکی است. افزون بر این، باید اثر خطاها بر روی کارایی الگوریتم‌های کوانتومی را هم اضافه کرد. تقویت کوانتومی می‌تواند احتمال موفقیت این الگوریتم‌ها در حضور خطا را به شدت افزایش دهد. از این رو، پیدا کردن روش‌های موثری برای کاهش هزینه‌های محاسباتی الگوریتم‌های تقویت‌شده کوانتومی اهمیت دارد.

1. M Mezard, A Montanari. *Information, physics, and computation*. Oxford University Press (2009).
2. CH Papadimitriou, K Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Corporation (1998).
3. S Kirkpatrick, Jr CD Gelatt, MP Vecchi. *Science*. **220**(4598) (1983) 671-80.
4. E Farhi, J Goldstone, S Gutmann, J Lapan, A Lundgren, D Preda. *Science*. **292**(5516) (2001) 472.
5. R Monasson, R Zecchina, S Kirkpatrick, B Selman, L Troyansky. *Nature*. **400**(6740) (1999) 133.
6. M Mézard, G Parisi, R Zecchina. *Science*. **297**(5582) (2002) 812.

7. A Braunstein, R Zecchina. *Physical review letters*. **96**(3) (2006) 030201.
8. HC Nguyen, R Zecchina, J Berg. *Advances in Physics*. **66**(3) (2017) 197.
9. G Carleo, I Cirac, K Cranmer, L Daudet, M Schuld, N Tishby, L Vogt-Maranto, L Zdeborová. **91**(4) (2019) 045002.
10. M Gabrié, S Ganguli, C Lucibello, R Zecchina. *Neural networks: From the perceptron to deep nets. In Spin Glass Theory and Far Beyond: Replica Symmetry Breaking After 40 Years* 2023 (pp. 477-497).
11. F Krzakala, L Zdeborová. *Statistical physics methods in optimization and machine learning. Lecture Notes*. (2024).
12. L Zdeborová, F Krzakala. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*. **76**(3) (2007) 031131.
13. F Krzakala, A Montanari, F Ricci-Tersenghi, G Semerjian, L Zdeborová. *Proceedings of the National Academy of Sciences*. **104**(25) (2007) 10318-23.
14. M Mézard, G Parisi, MA Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company; (1987).
15. DH Wolpert, WG Macready. *IEEE transactions on evolutionary computation*. **1**(1) (2002) 67-82.
16. C Baldassi, A Ingrosso, C Lucibello, L Saglietti, R Zecchina. *Journal of Statistical Mechanics: Theory and Experiment*. **2016**(2) (2016) 023301.
17. F Pittorino, C Lucibello, C Feinauer, G Perugini, C Baldassi, E Demyanenko, R Zecchina. *Journal of Statistical Mechanics: Theory and Experiment*. **2021**(12) (2021) 124015.
18. C Gerbelot, E Troiani, F Mignacco, F Krzakala, L Zdeborova. *SIAM Journal on Mathematics of Data Science*. **6**(2) (2024) 400-27.
19. G Biroli, T Bonnaire, V De Bortoli, M Mézard. *Nature Communications*. **15**(1) (2024) 9957.
20. Y Dandi, L Pesce, L Zdeborová, F Krzakala. *arXiv preprint arXiv:2502.13961*. (2025).
21. C Baldassi, A Braunstein, A Ramezanpour, R Zecchina. *In Mathematical Foundations of Complex Networked Information Systems: Politecnico di Torino, Verrès, Italy 2009* (2015) (pp. 27-49). Cham: Springer International Publishing.
22. A Maleki. *Approximate message passing algorithms for compressed sensing* (Doctoral dissertation, Stanford University).
23. A Montanari, S Sen. *Foundations and Trends® in Machine Learning*. **17**(1) (2024) 1.
24. MJ Schuetz, JK Brubaker, HG Katzgraber. *Nature Machine Intelligence*. **4**(4) (2022) 367.
25. O Duranthon, L Zdeborová. *arXiv preprint arXiv:2503.01361*. (2025).
26. D Ghio, Y Dandi, F Krzakala, L Zdeborová. *Proceedings of the National Academy of Sciences*. **121**(27) (2024) e2311810121.
27. S Aranguri, G Biroli, M Mezard, E Vanden-Eijnden. *arXiv preprint arXiv:2501.00988*. (2025).
28. A Abbas, A Ambainis, B Augustino, A Bärtschi, H Buhrman, C Coffrin, G Cortiana, V Dunjko, DJ Egger, BG Elmegreen, N Franco. *Nature Reviews Physics*. (2024) 28:1-8.
29. LK Grover. *Physical Review Letters*. **95**(15) (2005) 150501.
30. A Ambainis. *ACM Sigact News*. **35**(2) (2004) 22.
31. PR Giri, VE Korepin. *Quantum Information Processing*. **16**(12) (2017) 315.
32. E Farhi, J Goldstone, S Gutmann. *arXiv preprint arXiv:1411.4028*. (2014).
33. L Zhou, ST Wang, S Choi, H Pichler, MD Lukin. *Physical Review X*. **10**(2) (2020) 021067.
34. K Blekos, D Brand, A Ceschini, CH Chou, RH Li, K Pandya, A Summer. *Physics Reports*. **1068** (2024) 1.
35. P Hauke, HG Katzgraber, W Lechner, H Nishimori, WD Oliver. *Reports on Progress in Physics*. **83**(5) (2020) 054401.
36. B Tasseff, T Albash, Z Morrell, M Vuffray, AY Lokhov, S Misra, C Coffrin. *Journal of Heuristics*. **30**(5) (2024) 325-58.
37. A Del Campo. *Physical review letters*. **111**(10) (2013) 100502.
38. JR Finžgar, Notarnicola S, Cain M, Lukin MD, Sels D. *arXiv preprint arXiv:2503.01958*. (2025).
39. A Ramezanpour. *Physical Review A*. **96**(5) (2017) 052307.
40. A Ramezanpour. *Physical Review A*. **106**(1) (2022) 012418.
41. S Lloyd, J J Slotine. *Physical Review A*. **62**(1) (2000) 012307.
42. J Preskill. *Fault-tolerant quantum computation*. Introduction to quantum computation and information. 213 (1998).
43. Z Cai, R Babbush, SC Benjamin, S Endo, WJ Huggins, Y Li, JR McClean, TE O'Brien. *Reviews of Modern Physics*. **95**(4) (2023) 045005.

44. A Ramezanpour. *J. Stat. Mech.* **123401** (2025).
45. D Shapira, S Mozes, O Biham. *Physical Review A*. **67**(4) (2003) 042301.
46. J Leng, F Yang, XB Wang. *Physical Review Research*. **7**(1) (2025) L012017.