

Dr hab. inż. Marcin Magdziarz
Katedra Matematyki Stosowanej
Politechnika Wrocławska
ul. Wybrzeże Wyspiańskiego 27,
50-370 Wrocław

Wrocław, 9 maja 2020 r.

Recenzja

rozprawy doktorskiej mgr. Szymona Majewskiego
pt. „Langevin Monte Carlo as convex optimization in the space of measures”

Uwagi ogólne. Rozprawa doktorska mgr. Szymona Majewskiego napisana została w języku angielskim, liczy 68 stron. Składa się z pięciu rozdziałów, dodatku oraz bibliografii liczącej 51 pozycji. W bibliografii znaleźć można jedną pracę doktoranta wspólną z promotorem dr. hab. Błażem Miasojedowem (UW) oraz promotorem pomocniczym dr. Alainem Durmusem (ENS Paris-Saclay). Praca ta została opublikowana w 2019 r. w wysoko notowanym czasopiśmie *Journal of Machine Learning Research* i została już zauważona przez innych badaczy – w bazie Google Scholar ma aż 32 cytowania.

Przedmiot rozprawy. W rozprawie analizowane jest klasyczne d -wymiarowe równanie Langevina

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dW_t, \quad (1)$$

wyprowadzone w podstawowej postaci przez Paula Langevina ponad sto lat temu. Langevin otrzymał to równanie 1908 roku, wykorzystując klasyczne zasady dynamiki newtonowskiej, na wiele lat przed formalizującymi je matematycznie pracami Itô. W powyższym równaniu W_t to proces Wienera, natomiast U to potencjał, w którym porusza się cząsteczka brownowska.

Równanie Langevina jest nieprzerwanie, od ponad stu lat, przedmiotem intensywnych badań wielu naukowców na świecie, głównie ze względu na bardzo szerokie zastosowania w modelowaniu układów fizycznych. Również ostatnich latach, dzięki rozwojowi metod eksperymentalnych typu single particle tracking, równanie (1) z niegaussowskim szumem oraz losowym czasem okazało się być ważnym modelem opisującym dynamikę układów złożonych. Inne bardzo aktualne kierunki rozwoju, w których wykorzystuje się (1) to metody bayesowskie dla zbiorów danych typu big data oraz powiązane z nimi metody

Monte Carlo. Tematyka rozprawy jest klasyczna, nie ulega jednak wątpliwości, że bardzo aktualna i ważna.

Przy odpowiednich założeniach na potencjał U , równanie (1) posiada rozwiązanie stacjonarne. Jednakże miara niezmiennicza μ w jawnej postaci znana jest w niewielu przypadkach. Do jej aproksymacji stosuje się zatem metody Monte Carlo oparte o schemat Eulera dla (1), postaci

$$X_{k+1} = X_k - \gamma_{k+1} \nabla U(X_k) + \sqrt{2\gamma_{k+1}} G_{k+1}. \quad (2)$$

Tutaj γ_k to wielkości kolejnych kroków w schemacie, natomiast G_k to niezależne standardowe wektory gaussowskie.

Istnieje bardzo bogata literatura dotycząca analizy błędów oraz szybkości zbieżności schematu (2). Omawiana rozprawa poświęcona jest tym właśnie zagadnieniom, jednak doktorant proponuje tu zupełnie nowe, oryginalne podejście, wykorzystujące metody optymalizacyjne na odpowiednich przestrzeniach metrycznych miar probabilistycznych (przestrzeń Wassersteina). Inspiracją dla autora jest praca Jordan et al. SIAM J. Math. Anal. 29 (1998), w której pokazano, że potok gradientowy funkcjonału energii związanego z potencjałem U jest rozwiązaniem w słabym sensie równania Kołmogorowa (Fokkera-Plancka) odpowiadającego (1). Dodatkowo, przy odpowiednich założeniach na wypukłość potencjału U , minimum funkcjonału energii jest równe mierze niezmienniczej μ . Powyższe zależności i analogie sugerują, że własności schematu aproksymacji (2) można badać za pomocą metod optymalizacji wypukłej na przestrzeniach miar probabilistycznych. W niniejszej rozprawie ta koncepcja została z sukcesem zrealizowana, co pozwoliło uzyskać szereg wyników, które poprawiają znane w literaturze wyniki innych autorów.

Chciałbym w tym miejscu podkreślić, że pomysł autora, aby do analizy równania Langevina i zbieżności schematu Eulera zastosować metody optymalizacyjne, jest bardzo oryginalny i nietrywialny. Od samego początku zrobił na mnie duże wrażenie.

Opis wyników rozprawy. Pierwszy rozdział stanowi wstęp, w którym zaprezentowana została tematyka rozprawy na tle istniejących w literaturze wyników.

W drugim rozdziale wprowadzono wykorzystywane w pracy definicje, twierdzenia i własności dotyczące równania Langevina, optymalizacji wypukłej, przestrzeni Wassersteina i funkcjonałów na nich, potoków gradientowych dla funkcjonałów energii oraz nierówności transportowych. Ze względu na skondensowaną formę, rozdział ten był trudny do czytania, ale niewątpliwie pomocny w zrozumieniu szczegółów późniejszych rozważań.

W rozdziale trzecim zamieszczone zostały najważniejsze wyniki rozprawy. Autor zaczyna od rozważań na temat analogii pomiędzy zagadnieniem wypukłej optymalizacji, a

wyznaczaniem miary niezmienniczej μ , która minimalizuje potok gradientowy funkcjonału energii. Rozważania te prowadzą do nierówności

$$2\gamma(\mathcal{F}_U(\mu R_\gamma) - \mathcal{F}_U(\nu)) \leq (1 - m\gamma)W_2^2(\mu, \nu) - W_2^2(\mu R_\gamma, \nu) + C\gamma^2, \quad (3)$$

która jest odpowiednikiem nierówności wykorzystywanej w jednokrokowych metodach optymalizacji wypukłej. Tutaj, γ – odpowiednio mała stała, $C > 0$, m – parametr silnej wypukłości U , ν – miara prawdopodobieństwa, $\mathcal{F}_U$ – funkcjonał energii, R_γ – jądro Markowa powiązane z (2), W_2^2 – metryka Wassersteina.

Kluczowa nierówność (3) została udowodniona w Tw. 3.1. Dowód polegał na osobnym szacowaniu ujemnej entropii oraz funkcjonału energii potencjalnej w trzech kolejnych Lematach 3.3 – 3.5. Zwłaszcza dowód Lematu 3.5, w którym wykorzystano słabe rozwiązania równania Fokkera-Plancka, wydał mi się ciekawy.

Z nierówności (3) autor wyprowadził szereg wyników dotyczących zbieżności schematu (2) do miary niezmienniczej μ . W szczególności, w Tw. 3.10 podano ograniczenie na odległość Wassersteina pomiędzy n -tą iteracją schematu Eulera, a miarą μ . Co istotne, wynik ten poprawia oszacowania dla stałych i niemalejących kroków γ uzyskane w pracach Dalalyan, Karagulyan (2019) oraz Durmus, Moulines (2019).

Rozdział czwarty stanowi rozszerzenie wyników rozdziału trzeciego na przypadek tzw. gradientu stochastycznego oraz subgradientu stochastycznego. W sytuacji, kiedy numeryczna ewaluacja gradientu ∇U jest kosztowna, w schemacie Eulera (2) stosuje się jego nieobciążony estymator. W wielu przypadkach pozwala to na istotne przyspieszenie obliczeń, dzięki czemu metody wykorzystujące gradient stochastyczny w ostatnich latach zyskały sporą popularność. Do analizy gradientów stochastycznych autor stosuje analogiczną strategię, jak w rozdziale trzecim. W Tw. 4.1 dowodzona jest jednokrokowa nierówność typu (3). Co ciekawe, wynik jest niemal identyczny, jak ten z rozdziału wcześniejszego. Oszacowanie różni się jedynie dodatkowym błędem rzędu γ^2 . W podobny sposób przebiega również sam dowód oraz analiza zbieżności schematu Eulera ze stochastycznym gradientem. Potrzebne jest tu jednak dodatkowe, naturalne założenie, że estymator gradientu ma jednostajnie ograniczoną wariancję. Na uwagę zasługuje fakt, że wyniki z tego rozdziału uzyskane zostały bez założenia o gładkości potencjału U , co jest rzadkością w literaturze.

Rozdział piąty poświęcony jest analizie odległości pomiędzy miarą niezmienniczą μ dla równania Langevina (1), a miarą niezmienniczą μ_γ dla procesu Markowa indukowanego przez schemat (2). Autor, wykorzystując Tw. 3.1 w prosty i elegancki sposób wyznacza w Tw. 5.2 oszacowania na względną entropię, całkowite wachanie oraz odległość Wassersteina pomiędzy μ oraz μ_γ . W dalszej części rozdziału, przy dodatkowych założeniach na regularność potencjału U , autor uzyskuje mocniejsze oszacowania na odległości pomiędzy

μ oraz μ_γ .

Znaczenie tematyki i ocena wyników. Główne wyniki rozprawy oraz pozostałe badane w niej zagadnienia dotyczące zbieżności schematu Eulera, są niewątpliwie aktualne i ciekawe. Na szczególne uznanie zasługuje pomysł autora, aby do analizy równania Langevina i zbieżności schematu Eulera zastosować metody optymalizacyjne. Jest to nowe i bardzo obiecujące podejście. Można mieć nadzieję, że z pomocą tego typu metod możliwe będzie również badanie równań Langevina z niegaussowskimi szumami, czy też ze zmienionym losowo czasem.

Badania w dziedzinie równań Langevina oraz ich rozszerzeń ze stochastycznym gradientem w dalszym ciągu są intensywnie rozwijana w wielu wiodących ośrodkach naukowych na świecie. Doktorantowi udało się dowieść szeregu oszacowań poprawiających znane dotąd wyniki uzyskane przez innych autorów. Dodatkowo, dzięki wykorzystaniu metod optymalizacyjnych, w przypadku gradientu stochastycznego udało się uniknąć założenia o gładkości potencjału U . Wyniki te niewątpliwie stanowią istotny wkład do teorii procesów stochastycznych, a w szczególności stochastycznych równań różniczkowych. Zostały one już zauważone przez wielu matematyków oraz fizyków.

Metody matematyczne wykorzystywane w rozprawie są nietrywialne i często trudne technicznie. W przeprowadzonych dowodach autor musiał wykazać się sprawnością rachunkową oraz znajomością teorii prawdopodobieństwa i procesów stochastycznych, metod optymalizacji wypukłej oraz analizy funkcjonalnej.

Rozprawa jest zredagowana przejrzysto i starannie. Poza kilkoma drobnymi literówkami, nie znalazłem istotnych błędów lub nieścisłości. W pracy zabrakło mi praktycznych zastosowań uzyskanych wyników. Mając na uwadze fakt, iż jej tematyka jest ściśle związana z zastosowaniami, jest to sprawa istotna. Zastosowania uzyskanych wyników w problemie bayesowskiej regresji logistycznej znalazłem jednak w artykule autora opublikowanym w Journal of Machine Learning Research 20 (2019), więc czuję się usatysfakcjonowany.

Konkluzja. Podsumowując stwierdzam, że rozprawa mgr. Szymona Majewskiego spełnia wszystkie warunki ustawowe i zwyczajowe stawiane rozprawom doktorskim i wnoszę o dopuszczenie jej do dalszych etapów postępowania doktorskiego.

Marcin Napdriś