ACADEMIA ROMÂNĂ
SCOSAAR

REZUMATUL TEZEI DE ABILITARE

Optimizare Stochastică

Domeniul de abilitare: Matematică

Ion Necoară

Rezumat

0.1 Contribuţiile acestei teze

Considerăm spaţiul Euclidian $\mathbb{R}^n$, înzestrat cu produsul scalar $\langle x, y\rangle = x^T y$ şi norma corespunzătoare $\|x\| = \sqrt{\langle x, x\rangle}$[1]. Problemele de optimizare considerate în aceasta teză sunt versiuni ale următoarei probleme stochastice, având funcţia obiectiv exprimată ca medie a doi termeni:

$$\min_{x \in \mathrm{dom} F} F(x) := \mathbf{E}\left[f(x,\xi) + g(x,\xi)\right], \tag{1}$$

unde ξ este o variabilă aleatoare peste un spaţiu de probabilitate $(\Omega, \mathcal{F}, \mathbf{P})$. De asemenea, funcţiile $f : \mathbb{R}^n \times \Omega \to \bar{\mathbb{R}}$ şi $g : \mathbb{R}^n \times \Omega \to \bar{\mathbb{R}}$ sunt proprii, inferior semicontinue şi convexe în primul argument. Incertitudinile din majoritatea aplicaţiilor practice de optimizare au determinat ca optimizarea stochastică să devină un instrument esenţial pentru multe domenii din matematica aplicată, cum ar fi învăţarea automată şi statistică, control şi procesarea semnalului, reţele de senzori şi altele. Principala caracteristică a problemei (1) este faptul că nu putem evalua (sub)gradientul $\nabla F(x)$ al funcţiei F într-un punct $x \in \mathrm{dom} F$, adică nu avem acces la subdiferenţiala $\partial F(x)$. Pe de altă parte, pentru o realizare dată $\hat{\xi}$ a variabilei aleatoare ξ, putem accesa $f(x,\hat{\xi}), g(x,\hat{\xi})$ şi respectiv (sub)gradienţii $\nabla f(x,\hat{\xi}), \nabla g(x,\hat{\xi})$. În această teză presupunem că (sub)gradienţii $\nabla f(x,\hat{\xi}), \nabla g(x,\hat{\xi})$ sunt estimatori imparţiali:

$$\mathbf{E}\left[\nabla f(x,\xi) + \nabla g(x,\xi)\right] \in \partial F(x).$$

De notat este faptul că această problema de optimizare (1) este foarte generală şi apare în multe aplicaţii din inginerie, statistică şi învaţare automată.

Un caz particular al problemei stochastice (1) este $g \equiv 0$ şi $F(x) = \mathbf{E}\left[f(x,\xi)\right]$. Aceasta este problema uzuală întâlnită în aplicaţiile de învaţare automată. În acest caz, cel mai utilizat algoritm pentru rezolvarea acestei probleme este stochastic gradient descent (SGD) propus de Robbins şi Monro[2]:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k, \xi_k),$$

unde ξ_k este o realizare a variabilei aleatoare ξ (adică un eşantion $\xi_k \sim \mathbf{P}$) şi pentru a garanta convergenţa acestui proces stochastic iterativ pasul α_k trebuie ales de forma

$$\alpha_k = \frac{\alpha_0}{k^\gamma}, \quad \text{unde } \alpha_0 > 0 \text{ si } \gamma \in [0,1].$$

În general, SGD converge lent şi de aceea accelerarea acestui algoritm este una din principalele direcţii de cercetare, mai ales în contextul actual de Big Data. Experimental, s-a observat că accelerarea convergenţei metodei SGD se poate realiza folosind minibatch (mai multe eşantioane):

$$x_{k+1} = x_k - \alpha_k \frac{1}{|J_k|} \sum_{\xi \in J_k} \nabla f(x_k, \xi),$$

[1]Majoritatea rezultatelor prezentate în această teză se pot extinde şi la spaţii mai generale.
[2]H. Robbins and S. Monro, A stochastic approximation method, The Annals of Mathematical Statistics, 1951.

unde mulţimea de eşantioane $J_k \subset \Omega$ are cardinalitatea $|J_k| > 1$. Însă, această iteraţie nu accelerează convergenţa către punctul de optim dincolo de o anumită valoare a dimensiunii minibatch-ului, aşa cum se observă din Figura 1[3]. În particular, pentru dimensiuni mari ale minibatch-ului avem nevoie de un număr mare de epoci (iteraţii), rezultând complexitate totală mare. Aşadar, o problemă deschisă în literatura de optimizare stochastică este identificarea condiţiilor pentru care avem garanţii matematice ca minibatch-ul accelerează convergenţa şi găsirea mărimii optime a minibatch-ului (numărul optim de eşantioane necesare la fiecare iteraţie).

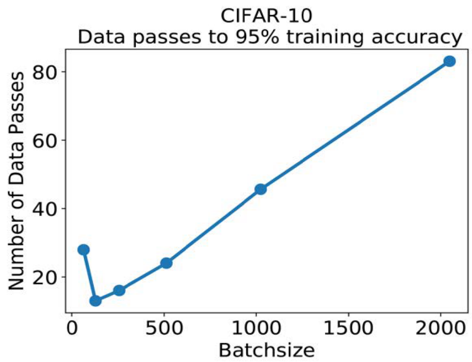

Figure 1: Compartamentul metodei SGD în funcţie de dimensiunea minibatch-ului.

În aceasta teză rezolvăm (parţial) această problemă, analizând versiuni minibatch al algoritmului Kaczmarz, în Capitolul 3 (amintim că algoritmul Kaczmarz coincide cu SGD atunci cand rezolvăm problema celor mai mici pătrate liniară), continuând cu algoritmi de proiecţie stochastică pentru problema de fezabilitate convexă, în Capitolul 4 şi în Capitolul 5 şi terminând cu un algoritm de tip stochastic subgradient pentru probleme de optimizare cu multe constrângeri, în Capitolul 7 .

În particular, în Capitolul 3, demonstrăm că algoritmul Kaczmarz, în care proiectăm simultan pe mai multe linii din matricea sistemului, converge liniar, având o rată de convergenţă care depinde de proprietăţile geometrice ale submatricilor sistemului liniar pe care proiectăm şi de dimensiunea lor (minibatch). Mai mult, analiza noastră de convergenţă arată că acest algoritm este eficient atunci când facem o eşantionare a liniilor matricei sistemului în submatrici bine condiţionate şi ne permite, de asemenea, să identificăm dimensiunea optimă a minibatch-ului.

În Capitolul 4, prezentăm o familie de metode de proiecţie stochastică pentru a rezolva probleme de fezabilitate convexă, exprimate în termeni de intersecţie (posibil infinită) de mulţimi convexe simple. Demonstrăm, sub anumite ipoteze de regularitate, că aceste metode au convergenţă liniară, cu o rată de convergenţă ce depinde de dimensiunea eşantionului de mulţimi considerat la fiecare iteraţie şi de o constantă ce are o interpretare naturală ca număr de condiţionare al problemei de fezabilitate. Am identificat constanta Lipschitz a problemei de fezabilitate ca parametru cheie care determină dacă minibatch-ul ajută la accelerarea convergenţei, precum şi cât de mult contribuie. În Capitolul 5, extindem aceste rezultate la probleme de fezabilitate convexă, unde fiecare mulţime din intersecţie este descrisă de o funcţie convexă Lipschitz. În acest caz, algoritmul nu presupune proiecţii pe mulţimi individuale, ci un pas de subgradient pentru una din funcţii. Din cunoştintele noastre, acestea sunt primele rezultate care determină teoretic condiţiile pentru care metodele de proiecţie stochastică de tip minibatch au o complexitate mai bună decât variantele lor bazate pe un singur eşantion.

[3]Preluată din Yin et al. 2018.

În Capitolul 7, considerăm probleme de optimizare cu multe constrângeri (posibil o infinitate) definite explicit cu ajutorul unor funcţii convexe Lipschitz. Pentru aceste probleme generale, am propus un algoritm stochastic de tip subgradient urmat de un pas de fezabilitate care utilizează doar subgradienţii unui eşantion de funcţii care definesc mulţimea fezabilă. La fiecare iteraţie, algoritmul face un pas de subgradient, menit să minimizeze funcţia obiectiv, şi apoi un pas de subgradient care minimizează încălcarea fezabilităţii eşantionului (minibatch-ului) de constrângeri observat. Derivăm rate de convergenţă liniară pentru acest algoritm care depind în mod explicit de dimensiunea minibatch-ul şi arătăm când minibatch-ul ajută o schemă de subgradient cu actualizări aleatorii de fezabilitate.

Pe de altă parte, analiza de convergenţă pentru SGD a fost dată pentru modele de optimizare simple, care satisfac ipoteze restrictive şi pentru care ratele de convergenţă sunt în general subliniare şi au loc doar pentru alegeri foarte specifice şi descrescătoare ale pasului. De exemplu, în literatura convexă, analiza de convergenţă a metodei SGD pentru cazul funcţiilor obiectiv Lipschitz este diferită de cea a funcţiilor cu gradient continuu Lipschitz, deşi ratele de convergenţă sunt la fel pentru ambele cazuri. De asemenea, analiza SGD-ului a fost derivată doar pentru cazul $g \equiv 0$. Însă, în multe aplicaţii practice apar constrângeri sau termeni de regularizare care conduc la o problemă de optimizare de forma (1), adică $g \not\equiv 0$. Extensia analizei convergenţei algoritmului SGD la cazuri mai generale, de exemplu pentru formularea (1), nu a fost facută încă. De aceea, în partea a doua a acestei teze, extindem analiza de convergenţă a algoritmului SGD pentru problema de optimizare generală (1).

Mai intâi, extindem analiza convergenţei metodei SGD la cazul problemelor de optimizare cu multe constrângeri exprimate, fie prin mulţimi convexe (Capitolul 6), fie prin funcţii convexe (Capitolul 7), posibil într-un număr infinit. În acest caz, g in (1) reprezintă funcţia indicator a unei mulţimi din familia de mulţimi care defineşte mulţimea fezabilă. De asemenea, deşi SGD este un algoritm uşor de implementat şi cu performanţe practice bune în anumite circumstanţe, studii recente arată că această metodă este instabilă şi sensibilă faţă de anumite alegeri ale parametrilor, de exemplu constanta de convexitate tare. De aceea, analizăm şi o versiune mai generală a metodei SGD, numită stochastic proximal point (SPP) în Capitolul 6. Din ratele de convergenţă derivate in Capitolul 6 pentru ambele cazuri, funcţii convexe Lipschitz şi cu gradient Lipschitz, se poate observa că SPP este, în general, atât mai robust la alegerea parametrilor, cât şi cu performanţe mai bune faţă de SGD.

În Capitolul 8, prezentăm un cadru teoretic pentru analiza metodelor stochastice de ordinul întâi (SGD şi SPP) pentru problema generală (1). Analiza noastră se bazează pe o ipoteză nouă asupra funcţiei obiectiv pe care am numit-o condiţia de mărginire a gradientului stochastic. Această condiţie include obişnuitele clase de funcţii analizate în literatură, funcţii Lipschitz şi funcţii cu gradient Lipschitz, şi ne permite să derivăm o analiză de convergenţă comună pentru metodele stochastice bazate pe (sub)gradienţi. În plus, ratele noastre de convergenţă sunt optime pentru această clasă de probleme.

În ultimul capitol prezentăm câteva direcţii de cercetare pe care dorim să le abordăm în viitor.

În final, menţionăm că toţi algoritmii prezentaţi în această teză au fost testaţi numeric pe aplicaţii concrete, folosind date sintetice sau reale, şi comparaţi cu alţi algoritmi din literatură. Rezultatele numerice, fie le confirmă pe cele teoretice, fie arată performanţe superioare ale metodelor noastre.

0.2 Articole în reviste ISI

Rezultatele prezentate în aceasta teză au fost publicate sau acceptate spre publicare în reviste ISI de top. Prezentăm mai jos lista de publicaţii din ultimii 3 ani, pe care se bazează această teză.

- **I. Necoara**, *Faster randomized block Kaczmarz algorithms*, Siam Journal on Matrix Analysis and Applications, vol. 40, nr. 4, 1425–1452, 2019 (Q1 if/ais - Applied Mathematics).
- **I. Necoara**, P. Richtarik, A. Patrascu, *Randomized projection methods for convex feasibility problems: Conditioning and convergence rates*, Siam Journal on Optimization, vol. 29, nr. 4, 2814–2852, 2019 (Q1 if/ais - Applied Mathematics).
- **I. Necoara**, A. Nedich, *Minibatch stochastic subgradient-based projection algorithms for solving convex inequalities*, partially accepted in Computational Optimization and Applications, 2020 (Q1 if/ais - Applied Mathematics).
- A. Nedich, **I. Necoara**, *Random minibatch subgradient algorithms for convex problems with functional constraints*, Applied Mathematics and Optimization, vol. 80, nr. 3, 801–833, 2019 (Q1 if/ais - Applied Mathematics).
- A. Patrascu, **I. Necoara**, *Nonasymptotic convergence of stochastic proximal point methods for constrained convex optimization*, Journal of Machine Learning Research, vol. 18, no. 198, 1–42, 2018 (Q1 if/ais - Automation & Control Systems).
- **I. Necoara**, *General convergence analysis of stochastic first order methods for composite optimization*, Journal of Optimization Theory and Applications, doi: 10.1007/s10957-021-01821-2, 2021 (Q2 if/ais - Applied Mathematics).

La două din lucrările de mai sus sunt unic autor, altele două au fost scrise în colaborare cu un fost student de la doctorat (A. Patrascu) şi două lucrări sunt în colaborare cu prof. A. Nedich de la Arizona State University (SUA).