Final Project for "Convex Optimization"

Zaiwen Wen
Beijing International Center for Mathematical Research
Peking University

November 18, 2019

1 Algorithms for large-scale Optimal Transport

1. Consider the standard form of LP

$$
\begin{aligned}
\min_{\pi \in \mathbb{R}^{m\times n}} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} \\
s.t. \quad & \sum_{j=1}^{n} \pi_{ij} = \mu_i, \quad \forall i = 1,\dots,m, \\
& \sum_{i=1}^{m} \pi_{ij} = \nu_i, \quad \forall j = 1,\dots,n, \\
& \pi_{ij} \geq 0.
\end{aligned} \tag{1.1}
$$

   (a) Solve (1.1) by calling mosek and gurobi directly in Matlab or python. The package "CVX" is not allowed to use here. Compare the performance between the simplex methods and interior point methods.

   (b) Write down and implement a first-order method, for example, the alternating direction method of multipliers.

   (c) Test problems:

   - Generate some random data c, μ and ν.
   - Find or construct the data sets in the references:
     - Jörn Schrieber, Dominic Schuhmacher, Carsten Gottschlich, DOTmark – A Benchmark for Discrete Optimal Transport.
     - Samuel Gerber, Mauro Maggioni, Multiscale Strategies for Computing Optimal Transport.

2. Read the reference:

   - Gabriel Peyre, Marco Cuturi, Computational Optimal Transport, `https://arxiv.org/abs/1803.00567`.
     some slides on optimal transport can be found at `https://optimaltransport.github.io/slides/`
   - Ernest K. Ryu, Yongxin Chen, Wuchen Li, Stanley Osher, Vector and Matrix Optimal Mass Transport: Theory, Algorithm, and Applications, `https://arxiv.org/abs/1712.10279`

(a) Find one of the most important optimization problems from the above references. Write down the background and formulation clearly.

(b) Write and implement an algorithm for the optimization problem in 2(a) from the chosen reference. Try to reproduce the numerical results in that reference.

(c) Try to write down and implement an algorithm covered in this course for the optimization problem in 2(a). This algorithm should be different from the one in 2(b).

3. Requirement:

(a) Compare the efficiency (cpu time) and accuracy (checking optimality condition) of different methods.

(b) Prepare a report including

- detailed answers to each question
- numerical results and their interpretation

(c) Pack all of your codes in one file named as "proj2-name-ID.zip" and send it to TA:
pkuopt@163.com

(d) If you get significant help from others on one routine, write down the source of references at the beginning of this routine.

2 Algorithms and analysis for shallow neural network

Let $f(\theta; x_i)$ be the output of a neural network and $\sigma(x)$ be one of the following element-wise activation function

$$\text{Relu}(x) = \max(0, x), \tag{2.1}$$

$$\text{Sigmoid}(x) = \frac{1}{1+\exp(-x)}, \tag{2.2}$$

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{2.3}$$

For a given data set $\{\{x_i, y_i\}_{i=1}^N\}$, consider the supervised learning problem:

$$\min_{\theta} \quad h(\theta) = \sum_{i=1}^{N} L(y_i, f(\theta; x_i)), \tag{2.4}$$

where L is the following loss functions

$$\ell_2 \text{ loss} : L(y_i, f(\theta; x_i)) = \frac{1}{2}\|y_i - f(\theta; x_i)\|_2^2. \tag{2.5}$$

In the following questions, the activation and loss function can be any of their combinations.

1. Consider the two-layer feed-forward neural network

$$f(\theta; x) := f_{W,v}(x) = \frac{1}{\sqrt{m}} v^\top \sigma(Wx),$$

where $x_i \in \mathbb{R}^d$ is the input data, $W \in \mathbb{R}^{m \times d}$, $v \in \mathbb{R}^m$ are weight matrice. The parameters are concatenated into one column vector $\theta = [\text{vec}(W)^\top, v^\top]^\top$, where $\text{vec}(W)$ transforms the matrix W into a vector by stacking all the columns of W one underneath the other.

(a) Compute the gradient (or subgradient) of $h(\theta)$ with respect to θ and estimate its Lipschitz constant.

(b) Compute the Hessian of $h(\theta)$ with respect to θ if it is available and estimate the upper and lower bound of its eigenvalues.

(c) Will the function $h(\theta)$ be strongly convex in a small neighbourhood of the global optimal solution? Either try numerical experiments on a few small examples or try to establish certain theoretical results.

(d) Suppose that **the gradient descent method** with certain line search schemes or or certain stepsize strategies is applied to solve (2.4). Write down the method and the corresponding convergence results from a few literatures or the classic textbooks on nonlinear programming such as

- "Numerical Optimization", Jorge Nocedal and Stephen Wright, Springer
- "Optimization Theory and Methods", Wenyu Sun, Ya-Xiang Yuan

Is it possible that this method converges to a global optimal solution of (2.4)? Either try numerical experiments on a few examples or try to establish certain theoretical results.

(e) Suppose that **the stochastic gradient method** is applied to solve (2.4). Write down the method and the corresponding convergence results from a few literatures. Is it possible that this method converges to a global optimal solution of (2.4)? Either try numerical experiments on a few examples or try to establish certain theoretical results.

(f) Suppose that **the KFAC method** is applied to solve (2.4)

- Optimizing Neural Networks with Kronecker-factored Approximate Curvature, James Martens, Roger Grosse, `https://arxiv.org/abs/1503.05671`

Write down the method and the corresponding convergence results from a few literatures. Is it possible that this method converges to a global optimal solution of (2.4)? Either try numerical experiments on a few examples or try to establish certain theoretical results.

2. Requirement:

(a) Test problems:

- random examples created by yourself.
- MNIST or Cifar-10.
- You can consider the experiment setup in the following paper:
  - Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks, Arora, Sanjeev, Simon S. Du, Wei Hu, Zhiyuan Li, and Ruosong Wang, `https://arxiv.org/abs/1901.08584`

(b) Requirements on numerical experiments:

- Specify the weight initialization schemes chosen in your experiments.
- Write all the prerequisites and the usage of your codes in detail.

(c) Prepare a report including:

- detailed answers to each question.
- numerical results and their interpretation if there are numerical experiments.

(d) Pack all of your codes in one file named as "proj2-name-ID.zip" and send it to TA: pkuopt@163.com

(e) If you get significant help from others on one routine, write down the source of references at the beginning of this routine.