

Рецензия на рукопись

«Ускоренные методы нулевого порядка в гладкой выпуклой стохастической оптимизации»

Хафизов Ф.А

Статья посвящена ускоренным методам нулевого порядка в гладкой выпуклой оптимизации. Заметно, что авторами проделана большая работа по получению научных результатов и их оформлению в виде статьи. Статья написана на актуальную тему и опирается на содержание свежих статей ведущих российских и зарубежных ученых. Содержание статьи имеет довольно логичную структуру: доказанные теоретически результаты подтверждаются численными экспериментами. Отмечу, что работа интересна в том числе тем, что используемый в ней математический аппарат использует в себе идеи и вычислительной математики, и теории оптимизации. Однако для того, чтобы помочь авторам в совершенствовании статьи, высажу некоторые замечания по работе:

| Введение | |
|---------------------|--|
| 1 | Подсекция 1.1. Возможно, стоит подробнее рассказать, при решении каких сложных задач могут быть использованы методы оптимизации, в том числе нулевого порядка |
| 2 | Подсекция 1.2. Возможно, стоит напомнить читателю, что такое стохастические градиентные методы и чем они отличаются от обычных градиентных методов. Не совсем понятен смысл третьего абзаца. Ускоренные методы могут быть стохастическими? В фразе «другую важную роль» не ясно, кто играет первую важную роль? |
| 3 | В подсекции 1.2. говорится про «метод аппроксимации градиента … для методов Франка-Вульфа…» и про «модификацию алгоритма Франка-Вульфа». Так в чем новизна работы? В новом способе аппроксимации градиента или в самом итерационном методе? Аналогичный вопрос возникает и при сравнении в численных экспериментах ускоренного градиентного спуска и обычных «GD» и «Nesterov». В частности, имеется ли доступ к градиенту в «GD» и «Nesterov»? Возможно, в виде псевдокода стоит вставить и алгоритмы для этих «обычных» методов. |
| 4 | Подсекция 1.2. Лишняя запятая перед «и сильно-выпуклой задач». |
| 5 | Подсекция 1.3. Опечатка в слове «Проводится» |
| Постановка задачи | |
| 6 | Подсекция 2.1. Лишняя запятая после «то есть». |
| 7 | Возможно, стоит и более явно вводить обозначения. Например, что такое $\delta(x)$? Явно стоит написать, что $i \sim U\{1, d\}$ – это дискретное равномерное распределение на отрезке (если я правильно понял). Формулы и математические термины стоит комментировать русским языком. Например «оракул возвращает <формула>» стоит переписать в виде «оракул возвращает зашумленное шумом значение f_δ функции f ». |
| 8 | В формуле (3) было бы здорово пояснить, зачем нужен множитель, равный размерности пространства d . |
| 9 | Что означает фраза «достаточно мало»? По сравнению с чем? |
| Основные результаты | |
| 10 | Является ли новым Accelerated Gradient Descent? Или из нового в нем лишь аппроксимация градиента? |
| 11 | Возможно, стоит один раз указать, что нормы векторов – это их 2-нормы. |
| 12 | Симметричные матрицы размерности d , вероятно, имеют обозначение S^d . Т.е. индекс |

| | |
|---------------|--|
| | размерности пространства – верхний. |
| 13 | Не совсем ясно, как вы нашли x^* . Задача ведь и заключается в ее поиске. То есть стоит подробнее расписать фразу «найденная численно». |
| 14 | Чем обусловлено такое большое число вызовов оракула? |
| 15 | Странно напечатано $\lambda = 0,1$. Между запятой и единицей большой пробел? |
| 16 | Возможно, стоит рис.1. и рис.2. поместить на одной высоте (в двух колонках). Так как сейчас рисунок 2 находится рядом с рисунками из серии других численных экспериментов. |
| 17 | Исходя из каких соображений выбирается метрика точности? Почему в разных экспериментах используются разные метрики? |
| Анализ ошибки | |
| 18 | В подсекции 5.2. вы пишете «за исключений изменений, описанных выше». Для улучшения читаемости работы, стоит их прописать еще раз. |

К коду замечаний нет.

Отмечу, что все указанные недочеты не несут критического характера. Их основной посыл – сообщить о местах в работе, которые могут быть непонятны не погруженному в тематику читателю.

Рецензент: Рубцов Д.Н.