Новый метод для оценки операторной нормы матрицы сверху

Ряполов Д. В.

Научный руководитель к.ф-м.н Рахуба М. В.

Национальный исследовательский университет «Высшая школа экономики»
Факультет компьютерных наук

15 мая 2025 г.

Содержание

1. Введение
2. Статистика Диксона
3. Адаптация степенного метода
4. Новый метод (Counterbalance)
5. Новый метод: практика
6. Адаптация Hutch++
7. Заключение

Постановка задачи

Оценка операторной нормы:

$$A \in \mathbb{R}^{n\times n},\ x \sim \mathbb{R}^n,$$

$$\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2.$$

Применения:

- Число обусловленности.
- Чувствительность матричных функций.
- Регуляризация нейросетей.

Ограничения

Общие:

- Sparse представление матрицы.
- Только матвек-операции $x \mapsto Ax$.
- $x \sim N(0, I_n)$.

Связанные со случайностью:

- Работа с гауссовскими квадратичными формами.
- Необходимо строгое теоретическое обоснование.

Статистика Диксона: теория

Статистика имеет вид

$$\mathbf{T}_\theta(\mathbf{A}, \mathbf{x}) = \theta(\mathbf{x}^\top \mathbf{A}^{\mathbf{m}} \mathbf{x}),$$

Известная теоретическая оценка

$$\mathbb{P}\big(T_\theta(A, x) \leq \|A\|_2\big) \leq \sqrt{\frac{2}{\pi}} \theta^{-m/2}.$$

Ключевые преимущества:

- Оценка является равномерной по A.
- При независимых x_i можно брать $\max_i T_\theta(A, x_i)$

Статистика Диксона: практика

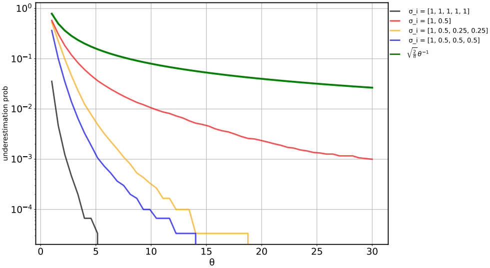

Рис. 2.1: классический метод — эмпирика vs. теор оценка

Ключевые недостатки:

- Оптимальное значение вероятности достигается на ранге 1.
- Оценка на полноранговых матрицах получается пессимистичной.
- При больших m (> 10) теряется преимущество в виде возможности параллелизации.

Адаптация степенного метода

Статистика:

$$\mathbf{P}_\theta(\mathbf{A}, \mathbf{x}) = \theta \, \frac{\|\mathbf{A}^\mathbf{T}\mathbf{A}\mathbf{x}\|_\mathbf{2}}{\|\mathbf{A}\mathbf{x}\|_\mathbf{2}}.$$

При работе над проектом получена теоретическая оценка.

Theorem (1)

Пусть $\rho = \frac{\sum_{i=2}^{k} \sigma_i^2}{\sigma_1^2}$, F *- кумулятивная функция. Тогда для* $\theta \geq 1$ *справедливо*

$$\mathbb{P}\left(P_\theta(A, x) \leq \|A\|\right) \leq F_{\chi_1^2}\left(\frac{\rho}{\theta^2 - 1}\right).$$

Зависимость оценки от ρ противоположна Диксону.

Новый метод «Counterbalance»

Объединяет классический и степенной:

$$B_\theta(x_1, x_2) = \sqrt{(P_\theta(A, x_1))^2 + (T_\theta(A, x_1))^2},$$

Theorem (2)

Пусть $\xi_1^2 \sim \xi_2^2 \sim N(0,1),\ \xi_1^2 \perp\!\!\!\perp \xi_2^2$. *Тогда справедливо*

$$\mathbb{P}\left(B_\theta(x_1, x_2) \leq \|A\|\right) \leq$$
$$\int_0^{\theta^{-2}} F_{\xi_1^2}\left(\frac{\rho t}{1-t}\right) p_{\xi_1^2 + \mathbb{I}[\rho \geq \theta^{-2}]\rho\xi_2^2}\left(\theta^{-2} - t\right) dt.$$

Новый метод: теория

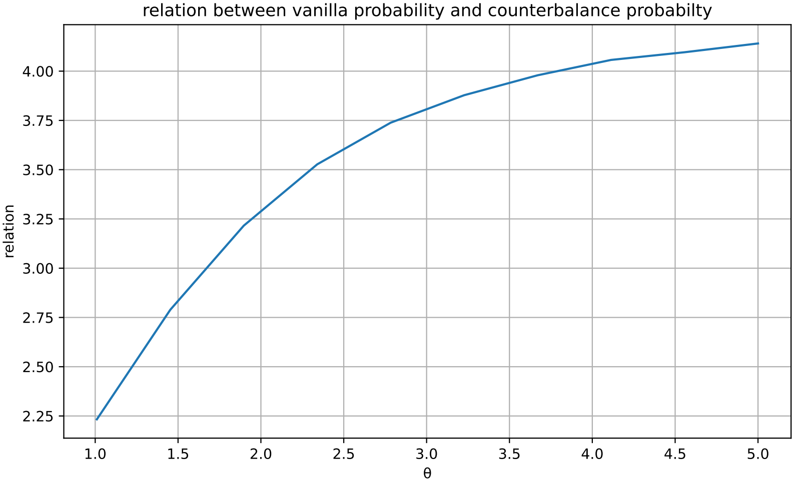

Рис. 5.1: отношение априорных вероятностей

Новый метод: практика

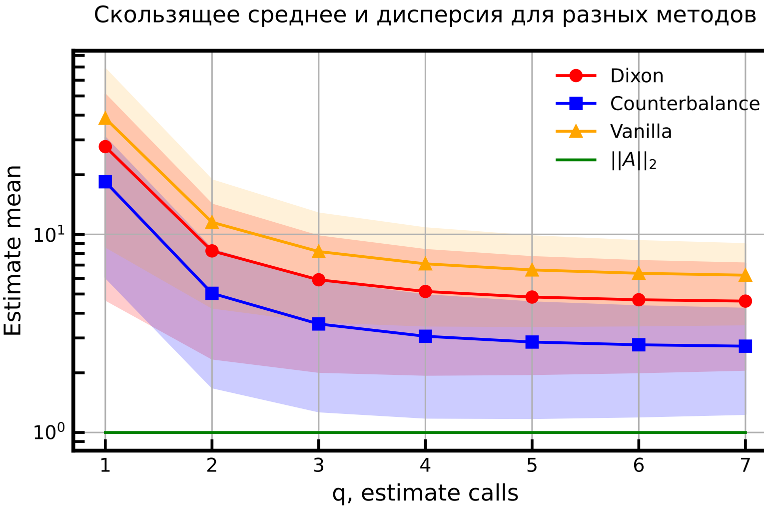

Рис. 5.1: отношение априорных вероятностей

Плотности оценок при $\theta = 5$, $p = 0.001$:

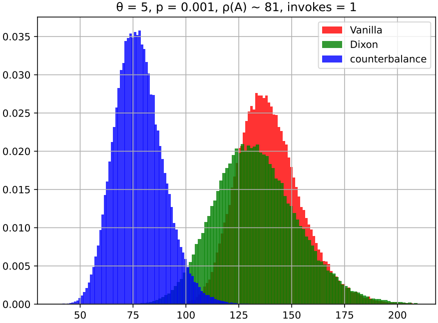

Рис. 5.2: сверточный слой

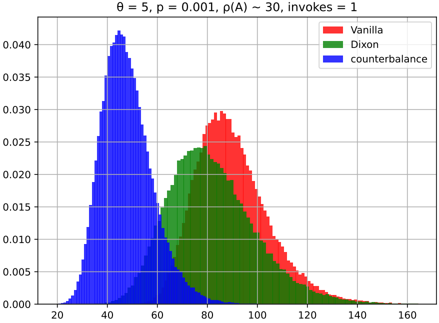

Рис. 5.3: производная Фреше

Адаптация Hutch++

Проектор $Q = \frac{Ax_1}{\|Ax_1\|}$. Статистика:

$$H_\theta(A, x_1, x_2) = \theta\sqrt{\|QQ^\top A\|_2^2 + \|(I - QQ^\top)Ax_2\|_2^2}.$$

Theorem (3)

Пусть $\mathrm{rk} A = 2$. *Тогда справедливо*

$$\mathbb{P}\left(H_\theta(A, x_1, x_2) \leq \|A\|_2\right) \leq \frac{0.1125}{\theta(\theta^2 - 1)}.$$

Адаптация Hutch++

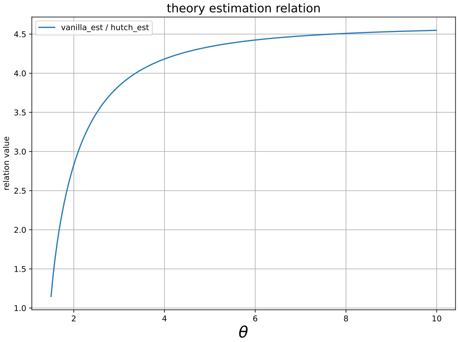

Рис. 5.4: Сравнение констант

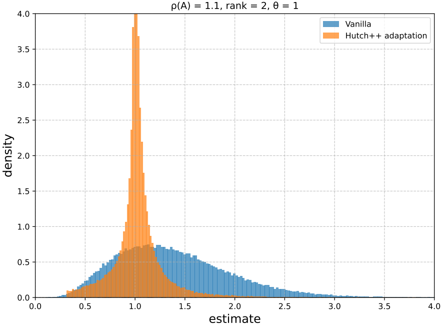

Рис. 5.5: эмпирическая плотность

Заключение и перспективы

Заключение

- Новые равномерно улучшают константу для априорной вероятности.
- Показана полезность методов на репрезентативных примерах.
- Предложен новый подход для преодоления тривиальных оптимальных случаев.

Перспективы

- Теоретическое обоснование адаптации Hutch++ для произвольного ранга.
- Расширение алгоритмов на онлайн формат.

References

[1] M.F. Hutchinson and. “A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines”. B: *Communications in Statistics - Simulation and Computation* 18.3 (1989), с. 1059—1076. DOI: 10.1080/03610918908812806. eprint: https://doi.org/10.1080/03610918908812806. URL: https://doi.org/10.1080/03610918908812806.

[2] Zvonimir Bujanovic и Daniel Kressner. “Norm and trace estimation with random rank-one vectors”. B: *SIAM Journal on Matrix Analysis and Applications* 42.1 (2021), с. 202—223.

[3] John Dixon. “Estimating extremal eigenvalues and condition numbers of matrices”. B: *Siam* (1983), с. 812—814.

[4] Henry Gouk и др. “Regularisation of neural networks by enforcing lipschitz continuity”. B: *Machine Learning* 110 (2021), с. 393—416.

[5] N. Halko, P. G. Martinsson и J. A. Tropp. “Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions”. B: *SIAM Review* 53.2 (2011), с. 217—288. DOI: 10.1137/090771806. eprint: https://doi.org/10.1137/090771806. URL: https://doi.org/10.1137/090771806.

[6] Judy Hoffman, Daniel A. Roberts и Sho Yaida. *Robust Learning with Jacobian Regularization*. 2019. arXiv: 1908.02729 [stat.ML].

[7] Raphael A. Meyer и др. “Hutch++: Optimal Stochastic Trace Estimation”. B: *2021 Symposium on Simplicity in Algorithms (SOSA)*, с. 142—155. DOI: 10.1137/1.9781611976496.16. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611976496.16. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611976496.16.

[8] Gabor J Szekely. “Extremal probabilities for Gaussian quadratic forms”. B: *Springer-Verlag* (2003).