

# Квазирандомизированный метод решения систем линейных уравнений<sup>1</sup>

С. И. Смирнов, аспирант<sup>2</sup>

Санкт-Петербургский государственный университет

q4ality@gmail.com

---

Как известно, решение задач о нахождении корня уравнения  $F(x) = 0$  и точки экстремума функции  $G(y)$ , когда значения функций  $F$  и  $G$  могут быть получены со случайными помехами, возможно с помощью методов стохастической аппроксимации.

Большой прикладной интерес представляет случай, когда упомянутые значения вычисляются с помощью методов Монте-Карло. В частности, если требуется решить большую систему линейных алгебраических уравнений, то вычислительную работу можно уменьшить за счет рандомизации или конструирования несмешенных оценок суммы квадратов невязок (случайный выбор  $k$  уравнений из  $n$ , где  $k \ll n$ ). В обоих случаях моделируется случайный процесс, решающий поставленную задачу. Представляет большой интерес использовать технику Квази-Монте-Карло метода (КМК) при моделировании.

В статье приведены алгоритмы стохастической аппроксимации для решения больших систем и обсуждаются результаты использования методов КМК для этих задач.

**Ключевые слова:** система линейных алгебраических уравнений, метод Монте-Карло, Квази-Монте-Карло метод, стохастическая аппроксимация.

## 1. Введение

Во многих областях прикладной математики требуется наличие инструмента, позволяющего с “достаточно хорошим” приближением вычислять результат умножения матрицы на матрицу, находить SVD-разложение или решение системы линейных алгебраических уравнений. При этом, зачастую, порядок матрицы для прикладных задач достаточно велик, что делает нецелесообразным применение стандартных итерационных методов (например, из [1, 2]) по причине их “достаточно большой” вычислительной сложности. В исследованиях по этой тематике предлагается использовать случайные несмешенные оценки матриц в стохастических итерационных алгоритмах, в том числе рассматривается возможность случайным образом выбирать лишь группу строк или столбцов матрицы и производить операции только над ними. Например, в работе [3] предполагается решать системы линейных алгебраических

---

<sup>1</sup>Работа выполнена при поддержке гранта РФФИ 11-01-00769-а.

<sup>2</sup>©С. И. Смирнов, 2011

уравнений с помощью схемы Неймана-Улама, моделируются оценки умножения матрицы на вектор, выбирая случайным образом подмножество столбцов матрицы. Аналогичные идеи можно найти в работах [4] (алгоритм умножения матриц) и [5] (для SVD). Как показано в этих и многих других работах, данный подход позволяет существенно уменьшить вычислительную сложность, но увеличивает погрешность аппроксимации.

В этой работе рассматривается комбинация подхода, основанного на рандомизированных оценках (идея описана выше), и метода Бочека, который используется как средство редукции размерности системы. Отметим, что в статье [6] обсуждается применение метода Бочека для систем линейных алгебраических уравнений, которые могут быть получены из метода Галеркина-Ритца. В данной работе подход естественным образом распространяется на более широкий класс матриц (предполагается только обратимость матрицы).

Хотелось бы отметить, что в работах [3–5] для построения процесса стохастической аппроксимации использовались только оценки, построенные с помощью метода Монте-Карло, тогда как в данной работе обсуждается возможность использования как МК, так и КМК.

В первой части работы формально описывается метод Бочека и используемый для расчетов процесс стохастической аппроксимации. Обсуждается вопрос вычислительной сложности.

Во второй части работы приводится обоснование сходимости данного метода стохастической аппроксимации (в терминах п. н.) и приводится аналог центральной теоремы для разницы между точным решением и оценкой с помощью метода стохастической аппроксимации.

В третьей части работы приводятся результаты моделирования как методом МК, так и КМК. Производится сопоставление результатов.

## 2. Постановка задачи. Описание метода стохастической аппроксимации

Рассмотрим систему линейных алгебраических уравнений, которая задается в виде:

$$AX = B, \quad (1)$$

где  $A$  — матрица размерности  $n \times n$ ,  $X$  — вектор неизвестных,  $B$  — правая часть, оба вектора из  $\mathbb{R}^n$ . В дальнейших рассуждениях будем считать, что система (1) имеет единственное решение (существует  $A^{-1}$ ).

Введем следующие обозначения:

$A(i, j)$  — элемент матрицы  $A$ , расположенный в  $i$ -ой строке и  $j$ -ом столбце,

$A(:, i)$  —  $i$ -ая строка матрицы  $A$ ,

$A(:, j)$  —  $j$ -ый столбец матрицы  $A$ ,

$\langle a, b \rangle$  — скалярное произведение векторов  $a$  и  $b$ ,

$$\|x\|_2 = \sqrt{\langle x, x \rangle}, \|x\|_1 = \sum_{i=1}^n |x(i)|.$$

**Определение 1.** Распределение  $P$  случайной величины  $\xi$ , принимающей значения на множестве  $1 : n$ , назовем согласованным с вектором  $Q \in \mathbb{R}^k$ , если для любого  $Q(i) \neq 0$ ,  $P(\xi = i) > 0$ .

Пусть имеются  $k$  линейно независимых векторов  $x_1, \dots, x_k$ ,  $k \ll n$ . Тогда в качестве приближения к решению задачи (1) рассмотрим такой вектор  $y = \sum_{i=1}^k C(i)x_i$ , который доставляет минимум функционалу  $\langle Ay - B, Ay - B \rangle$  среди всех векторов из линейной оболочки  $\{x_1, \dots, x_k\}$ . Так как значение скалярного произведения при заданных векторах

$x_1, \dots, x_k$  зависит только от  $C$ , то  $\langle Ay(C) - B, Ay(C) - B \rangle$  можно обозначить как  $g(C)$ . Задачу можно переписать в виде:

$$C_0 = \arg \min_{C \in \mathbb{R}^k} g(C). \quad (2)$$

Взяв частные производные по  $C(i)$  получим систему:

$$WC = V, \quad (3)$$

где  $U_i = Ax_i$ ,  $W(i, j) = \langle U_i, U_j \rangle$ , а  $V_i = \langle U_i, B \rangle$ .

Схема, описанная в работе [6], заключается в применении метода стохастической аппроксимации к системе (3).

## 2.1. Алгоритм стохастической аппроксимации для метода Бочека

Алгоритм стохастической аппроксимации является итерационным, поэтому достаточно показать переход из итерации  $N$  в итерацию  $N + 1$ , начальное значение — произвольный вектор  $C_1 \in \mathbb{R}^k$ ,

$\alpha$  — константа специального вида (параметр, обеспечивающий сходимость процесса для любой обратимой матрицы).

Переход от  $C_N$  к  $C_{N+1}$ .

1. Моделирование несмешенной оценки  $\widehat{W}$  матрицы  $W$ .

Для каждой пары  $(i, j) \in (1 : k) \times (1 : k)$  производится моделирование несмешенной оценки компоненты матрицы  $W(i, j)$  (компоненты матрицы моделируются независимо друг от друга):

1.1. Моделируются  $d$  независимых, равномерно распределенных на отрезке  $[0; 1]$  случайных величин.

1.2 С помощью равномерно распределенных величин получаются  $d$  реализаций  $\xi_1, \dots, \xi_d$  случайной величины, принимающей значение на множестве  $1 : n$  и имеющей распределение  $P(i, j)$ , согласованное с вектором  $Q \in \mathbb{R}^k$ , каждая компонента которого имеет вид:  $Q(l) = Ax_i(l) \cdot Ax_j(l)$ .

$$1.3. \widehat{W}(i, j) = \frac{n}{d} \cdot \sum_{s=1}^d Ax_i(\xi_s) \cdot Ax_j(\xi_s).$$

2. Моделирование несмешенной оценки  $\widehat{V}$  вектора  $V$ .

Для каждого  $i \in 1 : k$  производится моделирование несмешенной оценки  $V_i$ .

2.1. Моделируются  $d$  независимых, равномерно распределенных на отрезке  $[0; 1]$  случайных величин.

2.2 С помощью равномерно распределенных величин получаются  $d_1$  реализаций  $\eta_1, \dots, \eta_{d_1}$  случайной величины, принимающей значение на множестве  $1 : n$  и имеющей распределение  $P(i)$ , согласованное с вектором  $Q \in \mathbb{R}^k$ , каждая компонента которого имеет вид:  $Q(l) = Ax_i(l) \cdot B(l)$ .

$$2.3. \widehat{V}(i) = \frac{n}{d_1} \cdot \sum_{s=1}^{d_1} Ax_i(\eta_s) \cdot B(\eta_s).$$

3. Рассчитывается значение  $C_{N+1}$  по формуле:  $C_{N+1} = C_N - \frac{\alpha}{N} \cdot (\widehat{W}C_N - \widehat{V})$ .

Пусть было рассчитано  $N$  итераций алгоритма и любой вектор  $x_i$  имеет не более  $\lceil \frac{n}{k} \rceil + 1$  ненулевых компонент. Тогда очевидно, что было произведено порядка  $N \cdot O(nk + k^3)$  операций, что дает оценку вычислительной сложности повторения  $N$  итераций алгоритма.

### 3. Сходимость метода к решению задачи (3). Аналог центральной предельной теоремы

Пусть  $W_i$  — оценка для матрицы  $W$ , полученная на  $i$ -ом шаге алгоритма,  $V_i$  — оценка для  $V$ . Тогда верна следующая теорема (Теорема 1 из работы [9]):

**Т е о р е м а 1.** *Процесс стохастической аппроксимации (с итерацией  $C_{N+1} = C_N - \frac{\alpha}{N}(W_{N+1}C_N - V_{N+1})$ ) для симметричной, положительно определенной матрицы  $W$  сходится почти наверное к решению задачи (3) при любом начальном значении  $C_1$ , если оценки матрицы  $W$  и вектора  $V$  обладают следующими свойствами:*

1.  $\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{i=1}^N V_i - V \right\| = 0$  n. н.
2.  $\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{i=1}^N W_i - W \right\| = 0$  n. н.
3.  $\exists \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \|W_i\|^2$  n. н.

Отметим, что алгоритм стохастической аппроксимации, описанный в предыдущем разделе, удовлетворяет условиям этой теоремы, так как в силу согласованности распределения оценок с соответствующими векторами моменты оценок конечны, а по независимости и одинаковой распределенности оценок к ним можно применять усиленный закон больших чисел.

**Т е о р е м а 2.** *Пусть дана система  $WX = F$ , где  $W$  — положительно определена и симметрична,  $X_0$  — ее точное решение, и задана последовательность  $\{W_N, V_N\}$  независимых несмещенных оценок пары  $\{W, V\}$ , которая обладает следующими свойствами:*

1.  $W_N$  и  $V_N$  — независимы.
2.  $E(W_N(i,j)^4) < \sigma, \forall (i,j) \in (1:k) \times (1:k)$ .
3.  $E(V_N(i)^4) < \sigma, \forall i \in 1:k$ .

Если  $\alpha$ , удовлетворяет условию:  $2\alpha\lambda_{\min}(W) > 1$ , где  $\lambda_{\min}(W)$  — минимальное собственное число матрицы  $W$ , то процесс стохастической аппроксимации  $\{X_N\}$ , рекурсивно вычисляемый по формуле:

$$X_{N+1} = X_N + \frac{\alpha}{N} (W_{N+1} - V_{N+1}X_N),$$

где  $X_1$  — произвольный вектор, сходится почти наверное. В этом случае последовательность  $X_N$  обладает следующим свойством при  $N \rightarrow \infty$ :  $\sqrt{N}(X_N - X_0) \Rightarrow \mathcal{N}(0, S)$ , где  $\mathcal{N}(0, S)$  — случайный вектор, имеющий нормальное распределение с нулевым вектором средних и матрицей ковариаций

$$S = \int_0^{\infty} \alpha^2 e^{(\frac{I}{2} - \alpha W)t} S_0 e^{(\frac{I}{2} - \alpha W)t} dt,$$

в которой

$$S_0 = E \left( (W_N X_0 - V_N) \cdot (W_N X_0 - V_N)^T \right).$$

*Доказательство.* Эта теорема есть следствие теоремы 3.1 из работы [8]. Доказательство заключается в формальной проверке всех условий следствия 3.2 из теоремы 3.1 работы [8].  $\square$

Процесс стохастической аппроксимации, построенный в предыдущем разделе, удовлетворяет условиям теоремы 2, так как оценки моделируются согласованными с соответствующими векторами.

В силу того, что для процесса стохастической аппроксимации, построенного ранее, выполнены обе приведенных выше теоремы, можно рассматривать данный алгоритм как параметрически разделимый (если придерживаться терминов работы [7]), что позволяет распараллелить вычисления на любое произвольное число процессоров (подробнее о свойствах параметрически разделимых алгоритмов смотри [7]).

#### 4. Результаты моделирования. Вопрос о целесообразности применения КМК

Хорошо известен тот факт, что в задаче нахождения значения кратного интеграла применение КМК дает существенное улучшение (для МК норма ошибки имеет порядок  $O\left(\frac{1}{\sqrt{N}}\right)$ , а для КМК норма ошибки имеет порядок  $O\left(\frac{\ln^s N}{N}\right)$ ). Для того, чтобы исследовать улучшает ли КМК скорость сходимости процесса стохастической аппроксимации для данной реализации метода Бочека, были проведены следующие численные эксперименты при следующих упрощениях и допущениях:

1. Распределение вероятностей для моделирования оценок было взято равномерное.
2. Размер матрицы  $n$  равен  $m * k$ , где  $k$  — количество векторов  $x_1, \dots, x_k$ ,  $m$  выбиралась в зависимости от исследуемого случая.
3. Для нескольких комбинаций были промоделированы случайные матрицы соответствующих размерностей и “решения” системы, после чего был путем умножения получен вектор  $B$ .
4. Для каждого из получившихся уравнений были промоделированы процедуры стохастической аппроксимации, причем как с помощью метода МК, так и КМК.

5.  $d = 1$ .

Моделирование случайной матрицы:

Моделировались вектора  $p$  — каждая компонента которого была независимо, равномерно распределена на множестве  $[1, 6]$  и  $d$  — каждая компонента которого была равномерно распределена на  $[1, 5]$ . Из этих векторов были составлены диагонали матриц  $A_1$  и  $A_2$ . После чего для матрицы  $A_1$  были промоделированы только элементы под диагональю как независимые случайные величины с нормальным законом распределения, а элементы над диагональю были приравнены к нулю; а для матрицы  $A_2$  элементы над диагональю были получены как независимые реализации случайных величин с нормальным законом распределения, а элементы под диагональю были приравнены к нулю. Искомая матрица  $A$  получалась произведением матрицы  $A_1$  на  $A_2$ .

Точное решение задачи было получено следующим способом:

1. Как линейная комбинация векторов  $x_1, \dots, x_k$  со случайными коэффициентами, распределенными нормально со стандартным отклонением 5.
2. Как сумма компонент: аналогичной случаю 1 компоненты и случайной составляющей из  $q$  векторов, независимых, распределенных со стандартным нормальным распределением.

Выбор  $x_1, \dots, x_k$ :

Вектора были выбраны следующим образом:

$x_i(l) = 1, \forall l \in ((i-1)*m + 1 : i*m)$  и 0 для остальных компонент.

Оценки, моделируемые с помощью метода КМК, были получены при помощи последовательности Холтона с различными основаниями (в результатах приведены итоги для основания 3).

*Случай 1.* Вектора  $X$  моделировались в соответствии со случаем

$1, q = 0, n = 16, k = 4, m = 4$ . Было промоделированы три такие системы. Результаты приведены ниже ( $\text{error} = \langle C_0 - C_I, C_0 - C_I \rangle$ )

$Matrix_1, n = 16, k = 4, q = 0$				
$I$	$10^4$	$10^5$	$10^6$	$10^7$
$\text{error}_{MC}$	16.259	1.8568	0.1305	0.0654
$\text{error}_{QMC}$	12.6466	11.148	0.0994	0.0039

$Matrix_2, n = 16, k = 4, q = 0$				
$I$	$10^4$	$10^5$	$10^6$	$10^7$
$\text{error}_{MC}$	25.9293	14.0439	2.1724	0.1988
$\text{error}_{QMC}$	27.9051	10.6546	0.1250	0.0049

$Matrix_3, n = 16, k = 4, q = 0$				
$I$	$10^4$	$10^5$	$10^6$	$10^7$
$\text{error}_{MC}$	61.9445	8.1946	0.7892	0.08182
$\text{error}_{QMC}$	208.9331	35.4391	0.6287	0.0013

Для этих примеров подтверждается теоретический порядок сходимости метода МК (квадрат нормы ошибки имеет порядок  $O(N^{-1})$ ), тогда как скорость сходимости КМК заметно лучше. В то же самое время, улучшение скорости сходимости не столь значительно, как в случае вычисления кратного интеграла.

*Случай 2.* Вектора  $X$  моделировались в соответствии со случаем  $1, q = 0, n = 100, k = 10, m = 10$ . Было промоделированы две такие системы. Результаты приведены ниже ( $\text{error} = \langle C_0 - C_I, C_0 - C_I \rangle$ )

$Matrix_4, n = 100, k = 10, q = 0$				
$I$	$10^4$	$10^5$	$10^6$	$10^7$
$\text{error}_{MC}$	1987.1267	534.3881	65.0609	5.5982
$\text{error}_{QMC}$	1230.4543	300.4163	63.6743	1.1349

$Matrix_5, n = 100, k = 10, q = 0$				
$I$	$10^4$	$10^5$	$10^6$	$10^7$
$\text{error}_{MC}$	654.7632	314.7547	39.1172	5.0428
$\text{error}_{QMC}$	223.7816	64.9648	29.8086	0.9108

Ситуация аналогична случаю 1 но можно отметить, что алгоритм несколько дольше выходит на теоретический порядок сходимости.

*Случай 3.* Вектора  $X$  моделируются в соответствии со случаем 2,

$$\text{error} = \langle A * F * C_N - B, A * F * C_N - B \rangle,$$

где  $F$  – матрица, каждый столбец которой вектор  $x_i$ .

$Matrix_6, n = 64, k = 8, q = 10$				
$I$	$10^4$	$10^5$	$10^6$	$10^7$
$error_{QMC}$	$8.1473 \cdot 10^6$	$2.1473 \cdot 10^6$	$0.218 \cdot 10^5$	$3.1296 \cdot 10^2$
$error_{MC}$	$6.9723 \cdot 10^6$	$1.7654 \cdot 10^5$	$0.2944 \cdot 10^5$	$0.0357 \cdot 10^5$

$Matrix_7, n = 100, k = 10, q = 20$				
$I$	$10^5$	$10^6$	$10^7$	$2 * 10^7$
$error_{MC}$	83658650.9109	917135.9832	50849.8973	31656.3875
$error_{QMC}$	233007160.2677	3676261.0108	91784.8221	20778.1267

$Matrix_8, n = 128, k = 8, q = 50$			
$I$	$10^5$	$10^6$	$10^7$
$error_{QMC}$	$1.7177 \cdot 10^7$	$0.3581 \cdot 10^7$	$0.4423 \cdot 10^5$
$error_{MC}$	$2.6167 \cdot 10^6$	$1.3865 \cdot 10^6$	$1.0732 \cdot 10^5$

По представленным системам можно сделать вывод о том, что подтвердился порядок сходимости метода с оценками МК, и можно обоснованно выдвинуть гипотезу о том, что в случае использования в данном алгоритме оценок, моделируемых с помощью квазислучайных чисел, можно ожидать, что порядок сходимости будет лучше, чем  $O\left(\frac{1}{\sqrt{N}}\right)$ .

В целом, если моделировать оценки с меньшей дисперсией (один из инструментов уменьшения дисперсии – увеличивать  $d$ , кроме того, дисперсия существенно зависит от распределения вероятностей), то можно ожидать более быстрый переход к асимптотическому порядку сходимости, что делает в перспективе применение метода КМК более оправданным для данного класса алгоритмов стохастической аппроксимации (типа метода Бочека).

## Список литературы

- [1] Golub G.H., van Loan C.F. Matrix computations. — JHU. 1996. 3rd edition.
- [2] Грантмакер Ф.Р. Теория матриц – М.: Наука. 1966.
- [3] Sabelfeld K., Mozartova N. Sparsified Randomization Algorithms for large systems of linear equations and a new version of the Random Walk on Boundary method // Monte Carlo Methods Appl. 2009. Vol. 15. No. 3. P. 257–284.

- [4] *Drineas P., Kannan R.* Fast Monte Carlo algorithms for approximate matrix multiplication // Proc. of 42nd IEEE Symposium of Foundations of Computer Science. 2001. P. 452–459.
- [5] *Drineas P., Drinea E., Huggins P.S.* An experimental evalution of Monte Carlo algorithms for singular value decompositions // Lecture notes in computer science. 2003. Vol. 2563. P. 279–296.
- [6] *Bockek I.A.* Calculation of the Ritz and Galerkin coefficients by the Monte-Carlo method // USSR Computational Mathematics and Mathematical Physics. 1967. Vol. 7. No. 1. P. 231–239.
- [7] *Ермаков С. М.* Параметрически разделимые алгоритмы // Вестник СПбГУ. 2010. Сер. 1. Вып. 4. С. 25–32.
- [8] *Chen X., White H.* Asymptotic properties of some projection-based Robbins-Monro procedures in a Hilbert space // Studies in Nonlinear Dynamics and Econometrics. 2002. Vol. 6. P. 1–53.
- [9] *Györfi L.* Stochastic approximation from ergodic sample for linear regression // Z. Wahrscheinlichkeitstheorie verw. Gebiete. 1980. Vol. 54. P. 47–55.