



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 46 за 2024 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

А.А. Николаев, [С.В. Ершов](#),
[В.А. Фролов](#)

Интегрирование разрывной
функции значимости для
расчета освещенности
методом Монте-Карло с
динамикой Ланжевена

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](#)



Рекомендуемая форма библиографической ссылки: Николаев А.А., Ершов С.В., Фролов В.А. Интегрирование разрывной функции значимости для расчета освещенности методом Монте-Карло с динамикой Ланжевена // Препринты ИПМ им. М.В.Келдыша. 2024. № 46. 19 с.
<https://doi.org/10.20948/prepr-2024-46>
<https://library.keldysh.ru/preprint.asp?id=2024-46>

Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В. Келдыша
Российской академии наук

А.А. Николаев, С.В. Ершов, В.А. Фролов

Интегрирование разрывной функции
значимости для расчета освещенности
методом Монте-Карло с динамикой
Ланжевена

Москва — 2024

А.А. Николаев, С.В. Ершов, В.А. Фролов

Интегрирование разрывной функции значимости для расчета освещенности методом Монте-Карло с динамикой Ланжевена

Статья посвящена исследованию метода Монте-Карло на основе уравнения Ланжевена в применении к Монте-Карло рендерингу. В данной работе описана модельная сцена со сложным, разрывным ландшафтом функции значимости, в которой мы оцениваем влияние матрицы предобработки в уравнении Ланжевена. Такая сцена позволяет нам получить объяснимые и интерпретируемые результаты, что затруднительно в рендеринге на реальных 3D сценах. Результаты тестирования демонстрируют, что правильно построенная матрица предобработки позволяет значительно сократить число шагов метода Монте-Карло, необходимых для достижения заданной точности.

***Ключевые слова:** уравнение Ланжевена, методы Монте-Карло, MALA, алгоритм Метрополиса-Гастингса, цепи Маркова.*

A.A. Nikolaev, S.V. Ershov, V.A. Frolov

Langevin Monte Carlo integration with discontinuous contribution function for light transport simulation

Our paper is devoted to the study of the Monte Carlo method based on the Langevin equation as applied to Monte Carlo rendering. We describe a simplified scene with a complex and discontinuous density of contribution function, in which we evaluate the influence of the preconditioning matrix in the Langevin equation. Such a scene allows us to obtain explainable and interpretable results, which is difficult in real scenes encountered in Monte Carlo rendering. Test results demonstrate that a properly selected way to build preconditioning matrix can significantly reduce the number of steps of the Monte Carlo method, necessary to achieve the specified accuracy.

***Key words:** Langevin dynamics, Monte Carlo, MALA, Metropolis-Hastings method, Markov chain.*

1 Введение

При расчете освещенности в компьютерной графике и оптическом моделировании методы Монте-Карло, основанные на марковских цепях (МСМС, Markov Chain Monte Carlo), показывают хорошие результаты, когда интегрируемая функция имеет сложный рельеф. Среди различных вариаций МСМС следует отметить подходы, получившие название гибридного Монте-Карло [1, 2, 3, 4, 5, 6].

Они используют моделирование физических систем с целью улучшения сходимости МСМС, объединяя случайные блуждания с дополнительной информацией, старающейся учесть градиент и локальное поведение функции и, таким образом, формирующей анизотропные предложения перехода для алгоритма Метрополиса-Гастингса. Это приводит к улучшенной сходимости по сравнению с традиционными методами МСМС [7], делая их предпочтительным выбором для решения сложных вычислительных задач.

Однако в задачах компьютерной графики интегрируемая функция имеет ряд особенностей, которые не позволяют просто так перенести результаты исследований из области физического моделирования. В настоящий момент существуют всего две реализации в мире, успешно применившие гибридный Монте-Карло к задаче расчета освещенности трехмерных сцен [8, 9]. Причем обе реализации построены на базе одного программного кода и в действительности являются упрощенными методами Монте-Карло с использованием уравнений Гамильтона [8] и Ланжевена [9].

Нас интересовало, насколько допустимо для задач компьютерной графики делать те предположения, которые сделали авторы этих работ. Например, обе работы игнорируют разрывность интегрируемой функции, а матрица предобработки в работе [9] является диагональной и вычисляется только лишь на информации из градиента. Тем не менее в трехмерных сценах часто встречаются случаи, когда градиент интегрируемой функции близок к нулю (плоские поверхности с каустиками, рис. 1, освещенные участки стола на переднем плане), хотя при этом само интегрируемое пространство может быть довольно сложным.

Для исследования этой проблемы мы создали модельную сцену, на которой вычислительный эксперимент мог бы быть объясним и исследован с точки зрения параметров этой модельной сцены. Это позволяет нам получить интерпретируемые и объяснимые результаты, что затруднительно сделать для полноценного Монте-Карло рендеринга с применением марковских цепей и гибридного Монте-Карло.

2 Мотивация исследования

В данной работе мы говорим, что сходимость алгоритма улучшается или ускоряется, если уменьшается число итераций, необходимых для достижения фиксированной точности оценки. Считается, что в методе Метрополиса-Гастингса *число шагов*, необходимое для достижения заданной точности, ведет себя как $O(d)$, где d — размерность пространства интегрирования. Поскольку один шаг, очевидно, требует $O(d)$ операций, то *время счета* имеет асимптотику $O(d^2)$. Использование стохастической динамики Ланжевена для формирования предложения перехода в алгоритме Метрополиса-Гастингса позволяет улучшить эту асимптотику. Полученный таким образом алгоритм (MALA, Metropolis Adjusted Langevin Algorithm) имеет асимптотику числа шагов «всего лишь» $O(d^{1/3})$ ([10], теорема 1) и, таким образом, асимптотику времени счета $O(d^{4/3})$. Однако эти оценки получены для случая, когда интегрируемая функция гладкая. Вместе с тем, возникающая при трассировке лучей *функция значимости* (она же функция плотности) не всегда гладкая и, более того, может содержать частые разрывы (Рис. 1). Поэтому возникают два вопроса, требующих изучения:

1. Как ведет себя выгода MALA сравнительно с вариантом Метрополиса-Гастингса, использующим матрицу предобработки, при росте размерности задачи (в [11]–[13] размерность была фиксирована и невелика), и велика ли роль членов, специфичных для уравнения Ланжевена?
2. Как меняется соотношение выгоды разных вариантов метода MCMC, когда «направления максимального шага» имеют все большую кривизну (в [11]–[13] кривизна была фиксирована и невелика)?

К сожалению, непосредственное применение MALA для задачи расчета освещенности как на реальной, так и на модельной трехмерной сцене не позволит произвести анализ причин, по которым эффективность метода может падать, поскольку в обоих случаях марковская цепь исследует сложную многомерную функцию с неизвестным рельефом.

Поэтому мы исследуем поведение метода Монте-Карло с динамикой Ланжевена на функции специального вида — потенциальной яме. В терминах уравнения Ланжевена подынтегральное выражение $f(\mathbf{X})$ и потенциал $U(\mathbf{X})$ связаны соотношением $-\ln f(\mathbf{X}) = U(\mathbf{X})$. Вблизи потенциальной ямы $f \rightarrow 0+$ и $U \rightarrow -\infty$. В нашей тестовой сцене яма имеет вид «обобщенного тора», чье сечение есть $(N - 1)$ -мерный эллипсоид с постоянной формой и размерами, но чья ориентация вращается при обходе вокруг тора. Оказывается, что эта модельная задача для метода Монте-Карло с динамикой Ланжевена в действительности не такая простая, как кажется, а эксперименты с ней дают



Рис. 1: Пример 3D сцены, где в области больших значений интегрируемой функции наблюдается множество разрывов, обусловленных освещением через жалюзи. Изображение рассчитано с использованием Metropolis Light Transport в первичном пространстве путей.

нам возможность исследовать предельные случаи и построить анализ эффективности этого метода. В рассматриваемой модельной задаче есть несколько основных проблем, которые мы исследуем:

1. Эллипсоид рассеяния броуновского члена “протыкает” узкий овраг, что делает такое предложение перехода неэффективным. Мы предполагаем, что при увеличении размерности пространства эта проблема будет становиться все более острой;
2. Эллипсоид рассеяния броуновского члена при каждом шаге достаточно сильно поворачивается. Это приводит к тому, что шаг в обратном направлении становится маловероятным, а цепь осциллирует между границами тора, не попадая в его центральную часть.

В данной работе мы хотим ответить на вопрос: какие члены в уравнении Ланжевена помогают справиться с описанными выше проблемами и насколько хорошо. Кроме того, поскольку в приложениях расчета освещения вычисление градиента потенциала и дивергентного члена, как правило, является чрезвычайно дорогим, мы стремимся найти стратегию, основанную лишь на матрице предварительной обработки.

3 Langevin Monte Carlo

Далее для удобства будут использованы обозначения, принятые в [11]. Уравнение Ланжевена описывает движение частицы \mathbf{X} под воздействием случайных сил и сил вязкого трения с течением времени t :

$$d\mathbf{X} = -\nabla U dt + \sqrt{2dt}\boldsymbol{\xi}(t). \quad (1)$$

$\boldsymbol{\xi}(t)$ — случайный вектор, соответствующий случайной силе. Его компоненты независимы и имеют стандартное нормальное распределение. U — функция потенциала. Стационарное распределение полученного стохастического процесса описывается плотностью $\rho(\mathbf{X}) = Ze^{-U(\mathbf{X})}$ (Z — нормировочная константа, полученная из требования $\int \rho(\mathbf{X})d\mathbf{X} = 1$).

Для стохастического процесса \mathbf{X} нужно найти дискретную аппроксимацию \mathbf{X}_n . Будем использовать метод Эйлера-Маруямы, который для шага во времени δt определяет \mathbf{X}_n как

$$\mathbf{X}_n = \mathbf{X}_{n-1} - \nabla U(\mathbf{X})\delta t + \sqrt{2\delta t}\boldsymbol{\xi}(t). \quad (2)$$

Наша сцена для экспериментов будет содержать потенциальный овраг, вне которого плотность близка к нулю. Соответственно, для точек \mathbf{X} снаружи оврага $U(\mathbf{X}) = -\ln(\rho(\mathbf{X})) - \ln Z \rightarrow \infty$. Численное решение уравнения в текущем виде вынудит нас установить очень малый шаг по времени (иначе мы не достигнем достаточной точности), что повлечет за собой чрезмерно медленную сходимость.

Чтобы бороться с этим, в [1] предлагается сгладить рельеф U заменой переменных. Определим в каждой точке \mathbf{X} матрицу предобработки $\hat{T}(\mathbf{X})$ и будем использовать ее для замены переменных. Чтобы уменьшить анизотропию, $\hat{T}(\mathbf{X})$ должна “растягивать” пространство в областях с большим значением плотности и (наоборот) сжимать его, если плотность близка к нулю. После применения матрицы предобработки исходное уравнение примет вид

$$d\mathbf{X} = \mathbf{a}(\mathbf{X})dt + \sqrt{2dt}\hat{T}\boldsymbol{\xi}(t),$$

$$a_i(\mathbf{X}) = -\left(\hat{T}\hat{T}^*\nabla U\right)_i + \sum_j \frac{\partial(\hat{T}\hat{T}^*)_{i,j}}{\partial X_j}. \quad (3)$$

Следуя [11], будем называть (3) *обобщенным уравнением Ланжевена*. Оно содержит случайный броуновский член $\hat{T}\boldsymbol{\xi}(t)$ и сносовый член $\mathbf{a}(\mathbf{X})$. В свою очередь, $\mathbf{a}(\mathbf{X})$ содержит слагаемые $\sum_j \frac{\partial(\hat{T}\hat{T}^*)_{i,j}}{\partial X_j}$ — дивергентный член и $-\left(\hat{T}\hat{T}^*\nabla U\right)_i$ — градиентный член.

Аналогично (2) дискретизация обобщенного уравнения записывается как

$$\mathbf{X}_n = \mathbf{X}_{n-1} + \mathbf{a}(\mathbf{X})\delta t + \sqrt{2\delta t}\hat{T}\boldsymbol{\xi}(t). \quad (4)$$

Однако полученный дискретный процесс, вообще говоря, не будет сходиться к стационарному распределению исходного уравнения. Это вынуждает использовать правило Метрополиса-Гастингса как в [14]: очередную точку $\hat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ будем принимать только с вероятностью a . В случае непринятия состояние генерируемой последовательности не меняется. Вероятность a вычисляется как

$$a = \min\left(1, \frac{p(\mathbf{X}|\hat{\mathbf{X}})\rho(\hat{\mathbf{X}})}{p(\hat{\mathbf{X}}|\mathbf{X})\rho(\mathbf{X})}\right), \quad (5)$$

где ρ — плотность стационарного распределения $\rho(\mathbf{X}) = Ze^{-U(\mathbf{X})}$, $p(\hat{\mathbf{X}}|\mathbf{X})$ — плотность вероятности предложения перехода $\mathbf{X} \mapsto \hat{\mathbf{X}}$, а $p(\mathbf{X}|\hat{\mathbf{X}})$ — плотность вероятности предложения обратного перехода.

Из (4) видно, что плотность предложения имеет нормальное распределение $p(\mathbf{X}|\hat{\mathbf{X}}) = \mathcal{N}(\mathbf{X}|\mu, \Sigma)$ с соответствующими параметрами (средним μ и матрицей ковариации Σ):

$$\mu = \hat{\mathbf{X}} + \mathbf{a}(\hat{\mathbf{X}})\delta t, \quad \Sigma = 2\delta t \cdot \hat{T}(\hat{\mathbf{X}})\hat{T}(\hat{\mathbf{X}})^T.$$

Итоговая схема применения ЛМС имеет следующие шаги:

- Выбрать шаг по времени δt и способ вычисления матрицы предобработки $\hat{T}(\mathbf{X})$.
- На очередном шаге \mathbf{X}_n сгенерировать $\hat{\mathbf{X}}_{n+1}$.
- Рассмотреть $\hat{\mathbf{X}}_{n+1}$ как предложение перехода. Для этого нужно вычислить вероятность a в соответствии с (5). Для определения плотности $p(\mathbf{X}_n|\hat{\mathbf{X}}_{n+1})$ требуется значение матрицы предобработки и сносового члена в точке $\hat{\mathbf{X}}_{n+1}$, их тоже нужно вычислить.
- В случае успеха испытания Бернулли (с вероятностью a) положить $\mathbf{X}_{n+1} = \hat{\mathbf{X}}_{n+1}$. Иначе $\mathbf{X}_{n+1} = \mathbf{X}_n$.

Описанная схема также называется Metropolis-Adjusted Langevin Algorithm (MALA) [15, 9].

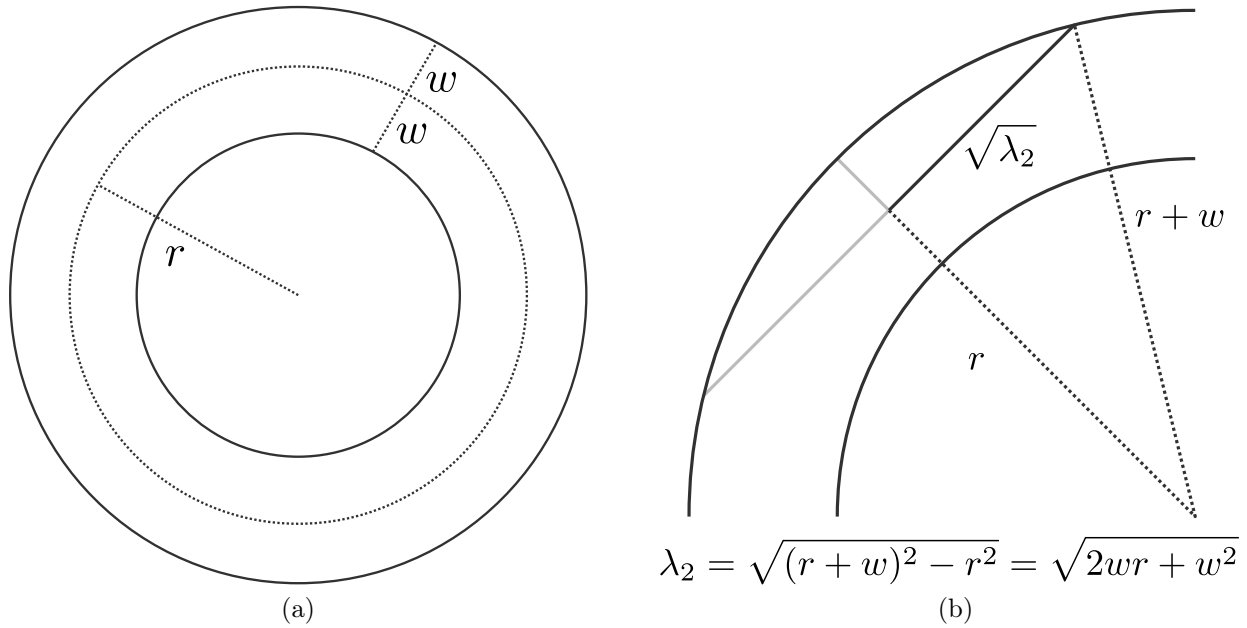


Рис. 2: Модельная сцена в двумерном пространстве имеет вид кольца радиуса r и ширины $2w$ (а).

Собственные значения λ предлагается вычислять как максимальные величины шага по радиальному и тангенциальному направлениям (б).

4 Двумерная модельная сцена

Рассмотрим плотность распределения двумерного случайного вектора. Создадим сцену, в которой плотность будет близка к нулю в точках, не находящихся в узком кольце. Область для такого кольца с радиусом $r > 0$ и полушириной $w > 0$ в координатах x, y задается уравнением $|\sqrt{x^2 + y^2} - r| \leq w$.

4.1 Функция плотности

В вычислительных экспериментах (раздел 7) мы используем функцию плотности, которая равна нулю вне области кольца, а внутри области принимает константное значение. То есть плотность $f_2(x, y)$ в двумерном случае определяется как

$$f_2(x, y) = \begin{cases} c_2, & |\sqrt{x^2 + y^2} - r| \leq w, \\ 0 & \text{иначе,} \end{cases}$$

c_2 — нормировочная константа.

4.2 Матрица предобработки

Матрицу предобработки в каждой точке определим так, чтобы область вытягивалась по тангенциальному направлению и сжималась по радиальному

направлению (то есть круг переходит в эллипс). При этом потребуем, чтобы собственные векторы матрицы соответствовали осям эллипса.

Обозначим собственные векторы как \mathbf{e}_i . Тогда $\hat{T}\hat{T}^* = \sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T$, где n — размерность пространства. Зная $\hat{T}\hat{T}^*$, мы можем вычислить \hat{T} либо аналитически, либо численно, используя разложение Холецкого.

Для $n = 2$ вектор \mathbf{e}_1 будет направлен от начала координат в сторону (x, y) , а \mathbf{e}_2 будет ему перпендикулярен:

$$\mathbf{e}_1 = \mathbf{e}_r = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{e}_2 = \mathbf{e}_\varphi = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} -y \\ x \end{pmatrix}.$$

Если рассматривать λ_1, λ_2 как величины шага вдоль радиального и тангенциального направлений соответственно, имеет смысл положить $\lambda_1 = w^2$, $\lambda_2 = (\sqrt{2wr - w^2})^2$ (Рис. 2). Однако такие значения не являются единственно возможными, потому далее будем считать $\lambda_1 = \lambda_1(w, r)$ и $\lambda_2 = \lambda_2(w, r)$.

В соответствии с (4.2), матрица с такими собственными векторами и собственными значениями

$$\hat{T}\hat{T}^* = \frac{1}{x^2 + y^2} \begin{pmatrix} \lambda_1 x^2 + \lambda_2 y^2 & (\lambda_1 - \lambda_2) xy \\ (\lambda_1 - \lambda_2) xy & \lambda_1 y^2 + \lambda_2 x^2 \end{pmatrix}. \quad (6)$$

В случае $n = 2$ относительно несложно аналитически найти \hat{T} . Для матрицы $\hat{A} = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$ верна формула:

$$\sqrt{\hat{A}} = \begin{pmatrix} \sqrt{A_{1,1} - A_{1,2}^2/A_{2,2}} & A_{1,2}/\sqrt{A_{2,2}} \\ 0 & \sqrt{A_{2,2}} \end{pmatrix}.$$

Применяя формулу к $\hat{T}\hat{T}^*$ из (6), получаем

$$\hat{T} = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} \sqrt{\lambda_1 x^2 + \lambda_2 y^2 - \frac{x^2 y^2 (\lambda_1 - \lambda_2)^2}{\lambda_2 x^2 + \lambda_1 y^2}} & \frac{xy(\lambda_1 - \lambda_2)}{\sqrt{\lambda_2 x^2 + \lambda_1 y^2}} \\ 0 & \sqrt{\lambda_2 x^2 + \lambda_1 y^2} \end{pmatrix}. \quad (7)$$

5 Трехмерная модельная сцена

Для трехмерного пространства с координатами x, y, z определим кольцо как множество точек с координатами, удовлетворяющими уравнению

$$\frac{(\sqrt{x^2 + y^2} - r_1)^2}{r_2^2} + \frac{z^2}{r_3^2} \leq 1,$$

где $r_i > 0$ и r_1 — радиус кольца, а r_2, r_3 — полуширины.

Рассмотрим усложненный вариант сцены: с изменением угла $\phi = \text{atan}_2(y, x)$ сечение трехмерного кольца будет равномерно вращаться (здесь и далее $\text{atan}_2(y, x)$ означает полярный угол точки (x, y) , т.е. $\text{atan}_2(y, x) = \text{Arg}(x + iy) - \pi$). В нашем эксперименте угол наклона эллипса $\theta = \theta(\phi) = k \cdot \phi$. Чтобы положение эллипса возвращалось в исходное при полном обороте ϕ , потребуем $k \in \mathbb{Z}$.

В данном разделе мы вводим два способа построения матрицы преобразования. Матрица, полученная первым способом, аналогична матрице в двумерной сцене и обозначается как \hat{T} . Во втором способе более детально исследуется траектория отдельно взятой точки. Это позволяет получить матрицу \hat{T}_2 , которая точнее подстраивается под локальный рельеф.

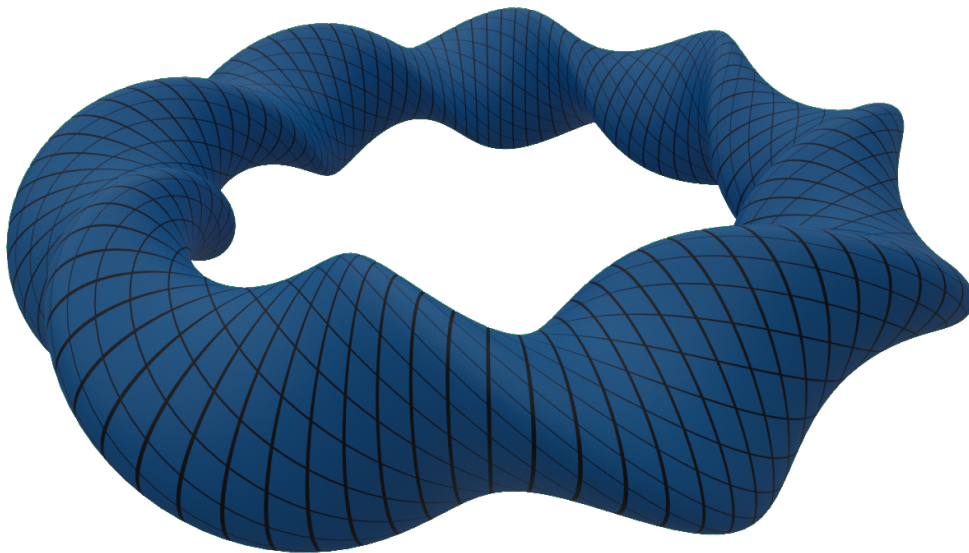


Рис. 3: Пример кольца в трехмерном пространстве с вращающимся сечением.

5.1 Вспомогательные системы координат

Для вывода формул для матриц преобразования мы будем работать с несколькими системами координат:

- Глобальная система x, y, z , центр кольца соответствует началу координат.
- Промежуточная система ϕ, ξ_r, ξ_z , представляющая собой сдвиг системы x, y, z так, чтобы начало координат находилось в точке $(x, y, 0)$.
- Локальная система $\phi, \xi_{\perp}, \xi_{\parallel}$, образованная поворотом ϕ, ξ_r, ξ_z так, чтобы оси координат соответствовали осям эллипса (сечения кольца).

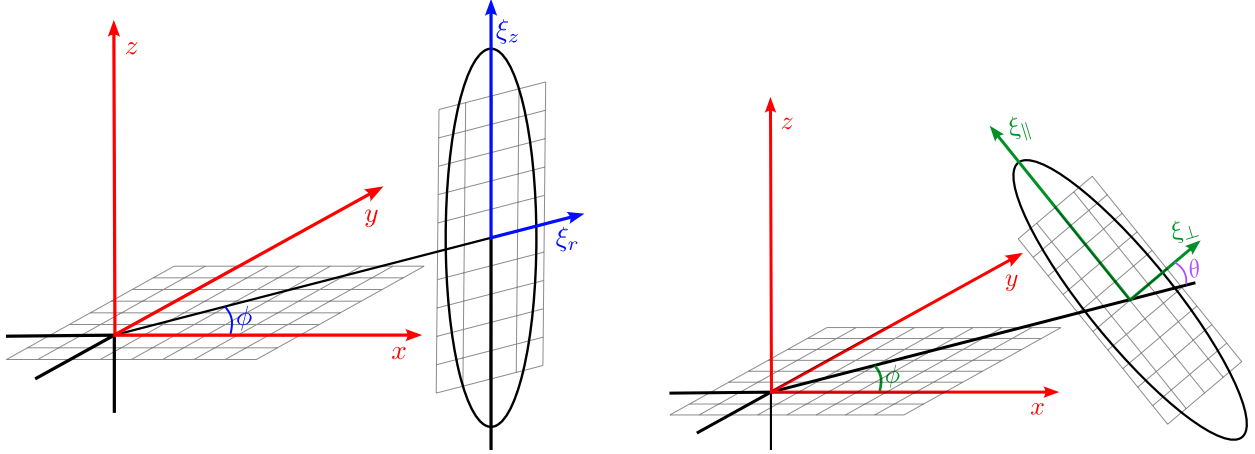


Рис. 4: Вспомогательные системы координат.

Координаты из промежуточной системы не участвуют в итоговых формулах. Далее будут важны только правила перехода между глобальной и локальной системами координат:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (R + \xi_{\perp} \cos \theta + \xi_{\parallel} \sin \theta) \cos \phi \\ (R + \xi_{\perp} \cos \theta + \xi_{\parallel} \sin \theta) \sin \phi \\ -\xi_{\perp} \sin \theta + \xi_{\parallel} \cos \theta \end{pmatrix}, \quad \begin{pmatrix} \phi \\ \xi_{\perp} \\ \xi_{\parallel} \end{pmatrix} = \begin{pmatrix} \operatorname{atan}_2(y, x) \\ \frac{x - R \cos \phi}{\cos \phi} \cos \theta - z \sin \theta \\ \frac{x - R \cos \phi}{\cos \phi} \sin \theta + z \cos \theta \end{pmatrix}.$$

5.2 Функция плотности

Теперь мы можем ввести f_3 — функцию плотности в трехмерной сцене:

$$f_3(\phi, \xi_{\perp}, \xi_{\parallel}) = \begin{cases} c_3, & \frac{\xi_{\perp}^2}{r_2^2} + \frac{\xi_{\parallel}^2}{r_3^2} \leq 1, \\ 0 & \text{иначе,} \end{cases}$$

c_3 — нормировочная константа.

5.3 Матрица предобработки \hat{T}

Ищем матрицу предобработки в следующем виде:

$$\hat{T}\hat{T}^* = \lambda_1 \cdot \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \cdot \mathbf{e}_2 \mathbf{e}_2^T + \lambda_3 \cdot \mathbf{e}_3 \mathbf{e}_3^T,$$

где \mathbf{e}_1 — вектор “вдоль направления ϕ ”, \mathbf{e}_2 — вектор “вдоль направления ξ_{\perp} ”, \mathbf{e}_3 — вектор “вдоль направления ξ_{\parallel} ”. При проведении вычислений мы положим $\sqrt{\lambda_1} = \sqrt{2r_1 \cdot \min(r_2, r_3)}$, $\sqrt{\lambda_2} = r_2$, $\sqrt{\lambda_3} = r_3$ (аналогично двумерному случаю).

Для нахождения \mathbf{e}_i умножим векторы, направленные вдоль осей координат x, y, z , на соответствующие матрицы поворота. Обозначив поворот вокруг

оси z как R_1 , поворот вокруг оси x как R_2 , а итоговый поворот как R , вычислим

$$R_1 := \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad R_2 := \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}, \quad R = R_1 \times R_2,$$

$$\mathbf{e}_1 = R \times [0, 1, 0]^T, \quad \mathbf{e}_2 = R \times [1, 0, 0]^T, \quad \mathbf{e}_3 = R \times [0, 0, 1]^T.$$

Учитывая написанное выше, справедливо

$$\hat{T}\hat{T}^* = R \times \begin{pmatrix} \lambda_2 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \times R^T.$$

Найти \hat{T} , как и ранее, можно численно или аналитически. Мы же будем использовать разложение Холецкого.

5.4 Матрица предобработки \hat{T}_2

Построенная в предыдущем разделе матрица \hat{T} будет плохо отвечать состоянию рельефа, когда $k > 0$. Действительно, рассмотрим траекторию точки $M(\phi) = M(\phi, \xi_{\perp}^0, \xi_{\parallel}^0)$ для $\phi \in [0, 2\pi]$. Направление движения M при каком-то ϕ не совпадает с направлением \mathbf{e}_1 .

Поэтому определим матрицу \hat{T}_2 , для которой \mathbf{e}_1 направлен как $\frac{\partial(x,y,z)}{\partial\phi}$. Теперь \mathbf{e}_1 , вообще говоря, не перпендикулярен \mathbf{e}_2 и \mathbf{e}_3 , потому \hat{T}_2 мы будем вычислять следующим образом:

$$\hat{T}_2 = \begin{pmatrix} \lambda_1 \cdot (\mathbf{e}_1)_x & \lambda_2 \cdot (\mathbf{e}_2)_x & \lambda_3 \cdot (\mathbf{e}_3)_x \\ \lambda_1 \cdot (\mathbf{e}_1)_y & \lambda_2 \cdot (\mathbf{e}_2)_y & \lambda_3 \cdot (\mathbf{e}_3)_y \\ \lambda_1 \cdot (\mathbf{e}_1)_z & \lambda_2 \cdot (\mathbf{e}_2)_z & \lambda_3 \cdot (\mathbf{e}_3)_z \end{pmatrix}.$$

Здесь λ_i — величины шага вдоль соответствующих направлений (как и для T).

6 Произвольная размерность

Переход от трехмерного случая (x, y, z) к четырехмерному (x_1, x_2, x_3, x_4) достаточно прост и прямолинеен. Сечением четырехмерного тора будет трехмерный эллипсоид, форму и размер которого мы по-прежнему полагаем постоянными. А вот его ориентация меняется при движении вдоль «осевой линии кольца», которая является окружностью и соответствует множеству центров эллипсоидов сечения. Ориентация определяется углом $\phi = \text{atan}_2(x_2, x_1)$.

У нашего эллипсоида три взаимно перпендикулярных оси $\mathbf{e}_2(\phi)$, $\mathbf{e}_3(\phi)$, $\mathbf{e}_4(\phi)$ (все они перпендикулярны

$$\mathbf{e}_1(\phi) = \frac{1}{\sqrt{x_1^2 + x_2^2}} \begin{pmatrix} -x_2 \\ x_1 \\ 0 \\ 0 \end{pmatrix},$$

лежащему по касательной к «осевой линии»), и эта тройка поворачивается относительно «невозмущенного» положения

$$\mathbf{e}_2(0) = \frac{1}{\sqrt{x_1^2 + x_2^2}} \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_4(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Несмотря на то что эта тройка направлений вложена в *четырёхмерное* пространство, ее положение может быть описано как поворот «трехмерной системы координат», который описывается, например, двумя углами Эйлера ϑ , φ . Очевидно, они должны быть (периодическими) функциями угла ϕ .

В N -мерном случае сечение будет $(N - 1)$ -мерным эллипсоидом. К сожалению, задание ориентации $(N - 1)$ взаимно перпендикулярных направлений с помощью $N - 2$ аналогов углов Эйлера невозможно. Попытки описания вращений в общем многомерном случае делались, и в литературе предлагаются несколько вариантов описания таких вращений с помощью углов, но они достаточно сложны и неудобны.

Для N -мерного случая мы зададим *какие-то* $N - 1$ векторов как функции угла ϕ , а затем ортонормализуем этот набор (сделаем его еще и ортогональным $\mathbf{e}_1(\phi)$). То, что получится, и возьмем в качестве $\mathbf{e}_2(\phi)$, $\mathbf{e}_3(\phi)$, \dots , $\mathbf{e}_N(\phi)$.

Например, взять для $i = 2, 3, \dots, N$

$$\mathbf{v}_i(\phi) = \left(\sum_{k \geq 1} \alpha_{i,j,k} \cos(mk\phi) \right) \mathbf{e}_i(0) + \sum_{j \neq i} \left(\sum_{k \geq 1} \beta_{i,j,k} \sin(mk\phi) \right) \mathbf{e}_j(0),$$

а векторы $\mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_N$ определить как

$$\mathbf{e}_2(\phi) = \frac{\mathbf{v}_2(\phi) - (\mathbf{v}_2(\phi) \cdot \mathbf{e}_1(\phi))\mathbf{e}_1(\phi)}{\|\mathbf{v}_2(\phi) - (\mathbf{v}_2(\phi) \cdot \mathbf{e}_1(\phi))\mathbf{e}_1(\phi)\|},$$

$$\mathbf{e}_3(\phi) = \frac{\mathbf{v}_3(\phi) - (\mathbf{v}_3(\phi) \cdot \mathbf{e}_1(\phi))\mathbf{e}_1(\phi) - (\mathbf{v}_3(\phi) \cdot \mathbf{e}_2(\phi))\mathbf{e}_2(\phi)}{\|\mathbf{v}_3(\phi) - (\mathbf{v}_3(\phi) \cdot \mathbf{e}_1(\phi))\mathbf{e}_1(\phi) - (\mathbf{v}_3(\phi) \cdot \mathbf{e}_2(\phi))\mathbf{e}_2(\phi)\|} \text{ и т.д.}$$

То, что в первый член представления $\mathbf{v}_i(\phi)$ входят только косинусы, а во второй — только синусы, имеет единственную цель обеспечить, чтобы $\mathbf{e}_2(0)$

совпадало с «невозмущенной ориентацией», что, в принципе, даже и необязательно. Множитель m для всех частот имеет тот же смысл, что и для трехмерного случая, а именно, увеличивая m , мы усиливаем кручение.

Коэффициенты $\alpha_{i,j,k}$ и $\beta_{i,j,k}$ формально произвольны. Однако их удачный выбор обеспечивает более равномерное и полное вращение (скажем, при слишком малых $\beta_{i,j,k}$ базис вообще поворачиваться не будет).

Как наиболее простой вариант, можно вообще взять

$$\mathbf{v}_i(\phi) = \cos(mk\phi)\mathbf{e}_i(0) + \sum_{j \neq i} \sin(mj\phi)\mathbf{e}_j(0), \quad i = 2, 3, \dots, N,$$

и даже это «заработает».

7 Результаты в двумерной сцене

В данном разделе мы рассмотрим поведение алгоритма в двумерном кольце при разных значениях w (r фиксировано и равно 1).

Так как коэффициент при броуновском слагаемом равен $\sqrt{2\delta t}$, для MALA без матрицы предобработки разумно установить шаг по времени так, чтобы выполнялось $\sqrt{\delta t} = w$, то есть $\delta t = w^2$. Для алгоритма с матрицей оптимальный δt будет другим, так как в расчетах матрицы уже учитываются параметры сцены. Он будет подобран так, чтобы доля принятых точек была сравнима с долей без матрицы.

Работы [11, 12, 13] показали, что сносовый член слабо влияет на скорость сходимости. В наших экспериментах он вычисляться не будет.

На графиках алгоритм с единичной матрицей обозначен как I, а с матрицей — как T.

Сравнение будет проводиться через вычисление среднеквадратичного отклонения по ячейкам двумерной гистограммы в координатах $\rho = \sqrt{x^2 + y^2}$, $\phi = \text{atan}_2(y, x)$ (10 ячеек по ρ , 50 по ϕ). Эталонное значение в ячейках было посчитано аналитически.

Широкое кольцо. Рассмотрим кольцо с относительно большим значением w : $w = 3 \cdot 10^{-1}$. На Рис. 5 (а) показаны результаты тестирования. Видно, что скорость сходимости не меняется при добавлении матрицы T .

Так как кольцо широкое, матрица со стандартными λ_1, λ_2 не может вытянуть области для генерируемых точек достаточно сильно, так что сходимость не улучшается.

Узкое кольцо. Теперь рассмотрим узкое кольцо с $w = 10^{-3}$. Результаты представлены на Рис. 5 (b). В такой сцене матрица T уже значительно ускоряет сходимость. После $\approx 10^{10}$ шагов у алгоритма с матрицей отклонение на два порядка меньше, чем у алгоритма без матрицы.

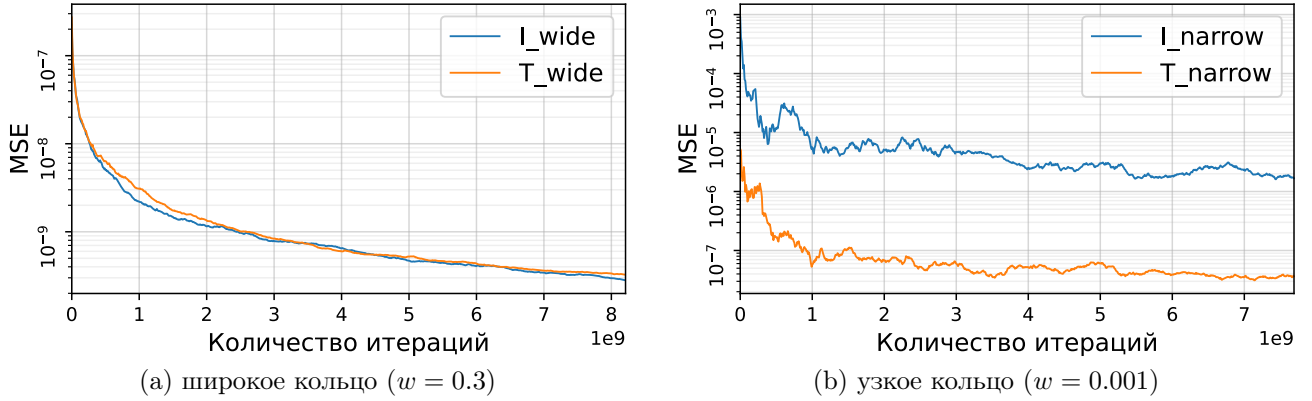


Рис. 5: Эксперименты с двумерным кольцом и гладкой функцией плотности: зависимость среднеквадратичного отклонения (MSE) от количества итераций для широкого (a) кольца и узкого (b).

8 Результаты в трехмерной сцене

Для сравнения будет строиться гистограмма на $(50, 20, 20)$ ячеек в координатах (ϕ, α, β) , где координаты α, β определяются следующим образом:

$$\alpha = \sqrt{\left(\frac{\xi_{\perp}}{r_2}\right)^2 + \left(\frac{\xi_{\parallel}}{r_3}\right)^2}, \alpha \in [0, 1],$$

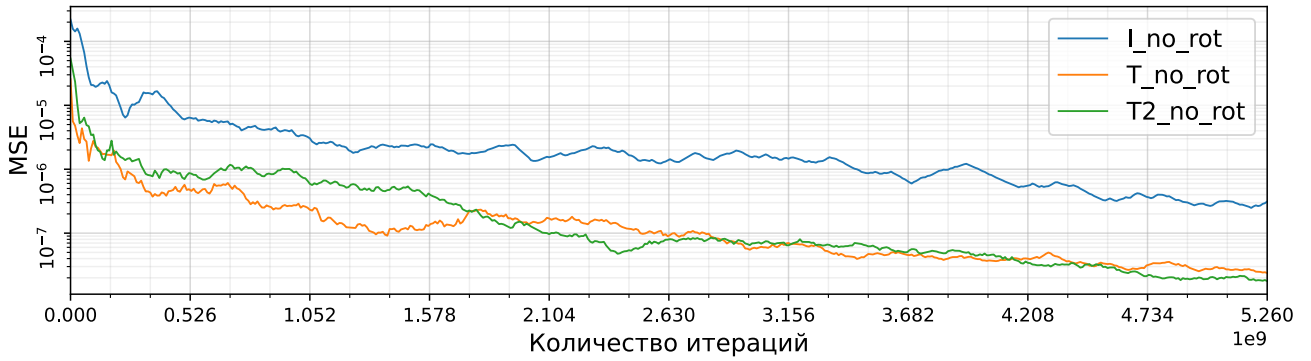
$$\beta = \text{atan}_2(\xi_{\parallel}, \xi_{\perp}), \beta \in [-\pi, \pi].$$

Как и в двумерном случае, мы отбрасываем дивергентный член, а r_1 фиксируем как $r_1 = 1$.

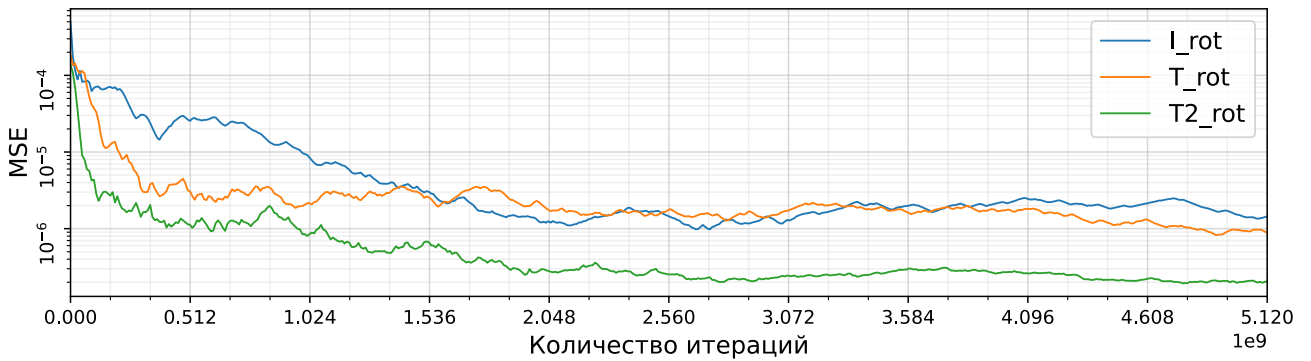
Для трехмерного кольца мы сразу перешли к “узкому” сечению, положив $r_2 = 10^{-3}$, $r_3 = 10^{-1}$. В этом разделе сравниваются результаты для трех алгоритмов с разными матрицами предобработки. На графиках единичная матрица и две матрицы T и T_2 обозначены как I , T и T_2 соответственно. На Рис. 6 суффикс `no_rot` соответствует сечению без вращения ($k = 0$), `rot` — с вращением ($k = 10$).

Матрица предобработки T . На Рис. 6 (a) показана работа алгоритмов с матрицей T и с единичной матрицей для кольца без вращающегося сечения. В такой сцене, как и в двумерном случае, матрица предобработки ускоряет сходимость.

Однако если положить $k \neq 0$, положительный эффект от матрицы T сходит на нет. На Рис. 6 (b) показаны результаты для кольца с вращающимся сечением при $k = 10$. В такой сцене на достаточно большом числе итераций алгоритмы с матрицами I и T показывают примерно одинаковую сходимость. Причины такого поведения были описаны в разделе 5.4 при выводе матрицы T_2 .



(a) кольцо без вращения ($k = 0$)



(b) кольцо с вращением ($k = 10$)

Рис. 6: Эксперименты с трехмерным кольцом: зависимость среднеквадратичного отклонения (MSE) от количества итераций для разных алгоритмов и степени закрученности (значения k).

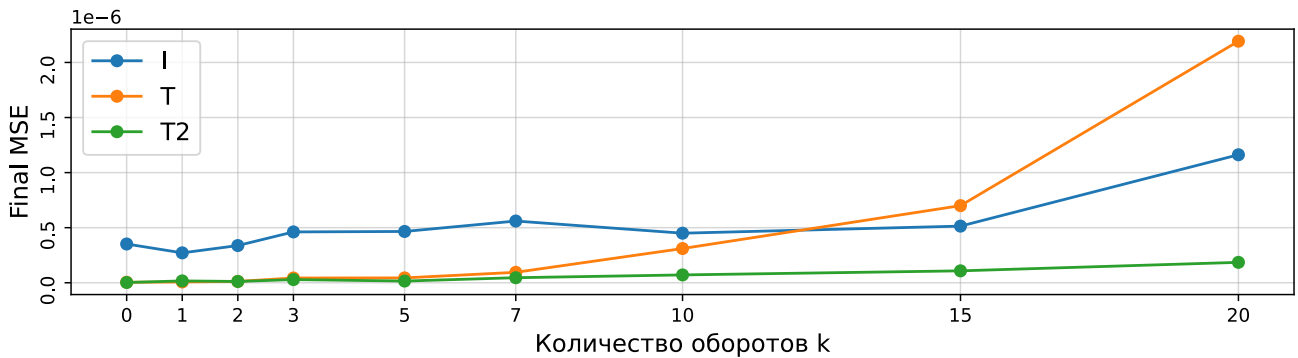


Рис. 7: MSE после $2 \cdot 10^9$ итераций для разных методов в зависимости от k . Так как кривая зависимости MSE от количества итераций содержит осцилляции, бралось не финальное MSE, а усреднение по 100 значениям MSE для количества шагов от 10^9 до $2 \cdot 10^9$.

Рис. 7 отражает зависимость среднеквадратичного отклонения после 10^9 итераций от k . Видно, что при увеличении k алгоритм с матрицей T начинает даже уступать алгоритму с единичной матрицей.

Матрица предобработки T_2 . Чтобы бороться с проблемой несоответствия матрицы T вращающемуся сечению, нами была предложена матрица

T_2 . При вращающемся сечении (Рис. 6 (b)) она показывает лучшие результаты, чем матрица T . Это значит, что T_2 действительно лучше подстраивается под изогнутый рельеф. Из Рис. 7 видно, что с увеличением k ошибка для T_2 увеличивается медленнее, чем для T .

При этом при $k = 0$ матрица T_2 переходит в T , поэтому на Рис. 6 (a) результаты для T и T_2 похожи.

9 Заключение

Мы показали, что для метода Метрополиса-Гастингса на основе динамики Ланжевена сходимость может быть значительно ускорена при использовании матрицы предобработки, подобранной под параметры сцены.

В рассмотренной модельной сцене в двумерном случае и трехмерном случае с $k = 0$ сходимость алгоритма с матрицей предобработки была лучше, чем с единичной матрицей. Однако не для всех сцен оптимальную матрицу предобработки легко подобрать. Трехмерное кольцо с вращающимся сечением демонстрирует эту проблему — матрица T , построенная по аналогии с двумерным случаем, не ускоряет сходимость. Матрица T_2 показывает себя существенно лучше, так как учитывает закрученность сечения. Однако построение такой матрицы в общем случае основано на дифференцировании, которое замедляет расчет и не всегда выполнимо. Мы прогнозируем, что как для многомерного кольца, так и для расчета освещенности методом Монте-Карло построение оптимальной матрицы предварительной обработки будет осложнено похожими проблемами. Таким образом, мы полагаем, что использование матрицы предварительной обработки для сцен с разрывной функцией значимости будет оправдано лишь в следующих случаях:

1. Если эту матрицу можно вычислять на основе просчитанной информации о разрывах в 3D сцене (например используя специальную фотонную карту, сохраняющую границы теней [16]);
2. Если эту матрицу можно приблизить блочно-диагональной формой, учитывая корреляцию лишь между несколькими соседними по номеру размерности переменными фазового пространства.

Это предмет наших будущих исследований.

Список литературы

- [1] Xifara T., Sherlock C., Livingstone S., Byrne S., Girolami M. Langevin diffusions and the Metropolis-adjusted Langevin algorithm // Statistics & Probability Letters. 2014. Т. 91. С. 14-19. arXiv preprint arXiv:1309.2983.

- [2] Schreck A., Fort G., Moulines E. Adaptive Equi-nergy Sampler: Convergence and Illustration // ACM Transactions on Modeling and Computer Simulation (TOMACS). 2013. Т. 23. № 1. Article 5. 27 pages. <https://doi.org/10.1145/2414416.2414421>
- [3] Braak C.J. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces // Statistics and Computing. 2006. Т. 16. № 3. pp. 239-249. <https://doi.org/10.1007/s11222-006-8769-1>
- [4] Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo // arXiv preprint arXiv:1701.02434. — 2017.
- [5] Homan M.D., Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo // J. Mach. Learn. Res. 2014. Т. 15. № 1. pp. 1593-1623.
- [6] Girolami M., Calderhead B., Chin S.A. Riemannian manifold hamiltonian monte carlo // arXiv preprint arXiv:0907.1100. — 2009.
- [7] Brooks S., Gelman A., Jones G., Meng X. Handbook of Markov Chain Monte Carlo. CRC press. 2011.
- [8] Li T.-M. Differentiable Visual Computing. — Ph.D. Thesis. MIT. 2019.
- [9] Luan F. Forward and inverse rendering with gradient based optimizations. — Ph.D. Thesis. Cornell University. 2021.
- [10] Sinho Chewi, Chen Lu, Kwangjun Ahn, Xiang Cheng, Thibaut Le Gouic, Philippe Rigollet. Optimal dimension dependence of the Metropolis-Adjusted Langevin Algorithm. arXiv preprint. <https://doi.org/10.48550/arXiv.2012.12810>
- [11] Ершов С.В., Фролов В.А., Николаев А.А., Волобой А.Г. Динамика Ланжевена в стохастической трассировке лучей: вычисление матрицы предобработки по ограничениям и выбор шага по времени // Препринты ИПМ им. М.В.Келдыша. 2023. № 63. 26 с. <https://doi.org/10.20948/prepr-2023-63>
- [12] Ершов С.В., Фролов В.А., Николаев А.А., Волобой А.Г. Динамика Ланжевена в стохастической трассировке лучей: выбор фазового пространства и ограничения на вариацию трассы // Препринты ИПМ им. М.В.Келдыша. 2023. № 64. 19 с. <https://doi.org/10.20948/prepr-2023-64>

- [13] Ершов С.В., Фролов В.А., Николаев А.А., Волобой А.Г. Динамика Ланжевена в стохастической трассировке лучей: вычислительные эксперименты // Препринты ИПМ им. М.В.Келдыша. 2023. № 65. 20 с. <https://doi.org/10.20948/prepr-2023-65>
- [14] Roberts G.O., Stramer O. Langevin diffusions and Metropolis-Hastings algorithms // Methodology and Computing in Applied Probability. 2002. Т. 4. № 4. pp. 337-357.
- [15] Luan F., Zhao S., Bala K., Gkioulekas I. Langevin Monte Carlo rendering with gradient-based adaptation // ACM Transactions on Graphics (TOG). 2020. Т. 39. № 4. Article 140. 16 pages.
- [16] Hua B. S. et al. Gradient-domain photon density estimation // Computer Graphics Forum. – 2017. – Т. 36. – №. 2. – pp. 31-38.