



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

Препринты ИПМ • Препринт № 63 за 2023 г.

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

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

Динамика Ланжевена в
стохастической трассировке
лучей: вычисление матрицы
предобработки по
ограничениям и выбор шага
по времени

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)



Рекомендуемая форма библиографической ссылки: Динамика Ланжевена в стохастической трассировке лучей: вычисление матрицы предобработки по ограничениям и выбор шага по времени / С.В. Ершов [и др.] // Препринты ИПМ им. М.В.Келдыша. 2023. № 63. 20 с.
<https://doi.org/10.20948/prepr-2023-63>
<https://library.keldysh.ru/preprint.asp?id=2023-63>

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В. Келдыша
Р о с с и й с к о й а к а д е м и и н а у к

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

Динамика Ланжевена в стохастической
трассировке лучей: вычисление матрицы
предобработки по ограничениям и выбор
шага по времени

Москва — 2023

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

Динамика Ланжевена в стохастической трассировке лучей: вычисление матрицы предобработки по ограничениям и выбор шага по времени

Основная вычислительно емкая задача реалистичной компьютерной графики — расчет глобальной освещенности. В работе проводится исследование скорости сходимости расчета освещения при помощи интегрирования методом Монте-Карло на основе уравнения Ланжевена. В статье представлена первая часть работы, в которой дается подробный обзор проблематики, исследуются вопросы влияния дивергентного члена, выбора шага интегрирования и выводятся формулы для вычисления матрицы предварительной обработки. Показано, как эти аспекты влияют на сходимость.

Ключевые слова: глобальная освещенность, стохастическая трассировка лучей, марковские цепи, уравнение Ланжевена.

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

Langevin dynamics in stochastic ray tracing: computing the preconditioning matrix according to restrictions and choice of time step

The main computationally extensive task of realistic computer graphics is the calculation of global illumination. The work investigates the speed of the convergence of lighting simulation using Monte Carlo integration based on the Langevin equation. The paper presents the first part of the work, which provides a detailed overview of the problem, examines the influence of the divergent term, the choice of the integration step, and derives formulae for calculating the preconditioning matrix. It is shown how these aspects affect convergence.

Key words: global illumination, stochastic ray tracing, Markov chain, Langevin equation.

1 Введение

Современная реалистичная компьютерная графика базируется на физически корректном моделировании распространения света в различных средах. Расчет глобальной освещенности, т.е. распределения света в виртуальной сцене, учитывающий множественные отражения и рассеяния света, всевозможные виды взаимодействия его с объектами сцены, является одной из основных вычислительно емких задач при построении реалистичного изображения. В основе расчета лежит уравнение рендеринга [1], ядром которого является интегрирование значений света, приходящего со всех возможных направлений. Существует несколько подходов к решению этой задачи, но в настоящее время большинство из них основываются на стохастической (Монте-Карло) трассировке лучей света.

В обзорной работе [2] рассматривается множество вариантов Монте-Карло интегрирования применительно к задаче расчета освещения. В ней делается вывод о преимуществах методов Монте-Карло на основе марковских цепей (МСМС — Markov Chain Monte Carlo). Эти методы можно рассматривать как расширение стандартного метода Монте-Карло. В то время как в обычном Монте-Карло выборки считаются независимыми, в МСМС они, наоборот, имеют корреляции между собой. Эта особенность МСМС позволяет эффективно использовать информацию об областях функции, имеющих высокую значимость. Один из наиболее распространенных вариантов МСМС — это алгоритм Метрополиса-Гастингса [3]. Основная цель всех алгоритмов МСМС заключается в построении распределения выборок, пропорционального произвольной целевой функции. Однако различные виды МСМС делают это по-разному, стремясь к эффективности предложенного решения [4].

В задачах компьютерной графики практическим результатом работы МСМС алгоритмов является накопленное среднее от фильтр-функции w_p по всем трассам лучей, проходящих через данный пиксель. Это среднее берётся по распределению, пропорциональному функции значимости I [5].

$$p \sim \langle w_p \rangle_I. \quad (1)$$

В работе [2] говорится о перспективности Монте-Карло методов на основе уравнений Гамильтона и Ланжевена, т.к. эти методы обладают лучшей сходимостью при росте размерности пространства интегрирования. Однако это методы новые, и многие их аспекты, включая вопросы эффективного применения, остаются неисследованными. Это и послужило мотивацией наших исследований, результаты которых изложены в трех препринтах [6–8].

2 Предыдущие работы

Традиционно [5, 9] трасса луча описывалась набором ее узлов (точек пересечения луча с поверхностями сцены) $\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_m\}$. Это представление называется иногда *мировым пространством путей*. Очевидно, можно использовать и другие параметризации трассы луча, например, через набор направлений ее сегментов $\mathbf{X} = \{\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_m\}$ [10]. В этом случае координаты узлов непосредственно не заданы, но вычисляются по $\{\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_m\}$. Возможны и иные представления, например, *первичное пространство путей*, в котором любая точка многомерного единичного куба соответствует некоторой трассе, возможно, правда экранированной (т.е. с нулевой функцией значимости). К этой группе методов относятся, например, Kelemen MLT [11] и MMLT [12].

При смене представления на $\mathbf{Y} = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_m\}$ плотность распределения трасс в новых переменных будет $I \left| \det \frac{\partial \mathbf{X}}{\partial \mathbf{Y}} \right|$; естественно полагать это функцией значимости в новых переменных. Тогда значения различных «измеряемых величин» можно по-прежнему вычислять как среднее по случайной выборке трасс (1), только теперь трассы надо генерировать с плотностью, пропорциональной уже *новой* функции значимости $I \left| \det \frac{\partial \mathbf{X}}{\partial \mathbf{Y}} \right|$. При удачном выборе параметризации \mathbf{Y} генерировать трассы с новым распределением может оказаться проще и эффективнее.

Один из наиболее прямолинейных путей выбора параметризации — это алгоритмы в так называемом *первичном пространстве путей* [12]. Второй способ — это работа в *мировом пространстве путей* (Path Space), в пространстве, составленном конкатенацией всех координат вершин путей в мировом пространстве (например, это методы Veach MLT [9], MEMLT [13, 14]). Третий способ заключается в комбинировании первых двух. В работах [15–17] строится обратимый переход между мировым пространством и первичным пространством путей, за счёт чего удастся объединить сильные стороны обоих подходов. Можно придумать много способов представления пространства интегрирования U . Практические аспекты методов расчёта освещения на основе МСМС хорошо представлены в [4]. Эта работа может служить в качестве отправной точки для понимания их на более глубоком уровне с целью практической реализации.

Таким образом, большинство существующих работ, применяющих МСМС к задаче расчета освещения, основаны на алгоритме Метрополиса-Гастингса или даже более простом алгоритме Метрополиса с равновероятным прямым и обратным переходом [4]. С другой стороны, развитие самих методов МСМС ушло далеко вперёд. Применение динамики Ланжевена в теории возможно к любому из представлений функции вклада. Работа, в которой это удалось сделать в настоящее время [18], основана на MMLT. В нашей работе мы используем пространство, идейно близкое к методу HSLT [10].

2.1 Работы на основе “гибридного” Монте-Карло

Гибридный метод Монте-Карло представляет собой обширный подкласс МСМС [19–24]. Эти методы осуществляют генерацию выборок через траектории динамической системы, такие как гамильтонова механика (НМС — Hamiltonian Monte Carlo) или броуновское движение в вязкой среде на основе уравнения Ланжевена (ЛМС — Langevin Monte Carlo). Основная идея этих методов состоит в том, что динамическая система позволяет генерировать удачные предложения перехода и хорошо обходит пространство, т.к. умеет двигаться в “нужных направлениях”.

Динамическая система Гамильтона используется в методе Hessian-Hamiltonian Monte Carlo (ННМС) [25, 26]. Особенностью метода является то, что дорогостоящее моделирование динамики Гамильтона производится не на самой функции потенциала, а на его квадратичном приближении через разложение в ряд Тейлора. Это существенный момент, поскольку уравнения Гамильтона для квадратичной функции имеют аналитическое решение. Далее, Hamiltonian Monte Carlo требует вычисления первой (градиент) и второй (гессиан) производных, которые были получены в работе при помощи библиотеки автоматического дифференцирования [27]. В центре метода ННМС лежит вычисление гессиана, который используется в алгоритме два раза. Первый раз — для вычисления аппроксимации потенциала через ряд Тейлора. Вторым раз — для inverse mass matrix (аналог матрицы предварительной обработки). При этом само предложение перехода делается в виде многомерной вытянутой гауссианы. Вычисление гессиана дорого, что является основным лимитирующим фактором метода.

В работе [28] был предложен относительно простой способ удешевления подходов, основанных на ННМС. В начале каждого шага делается обычное предложение перехода с использованием правила Метрополиса. Если оно принимается, алгоритм продолжает свою работу дальше как обычный шаг Метрополиса. Если же предложение перехода отвергается, новая попытка уже делается при помощи Hamiltonian Monte Carlo. Основная сложность в этом подходе состоит в корректном вычислении вероятностей предложения перехода.

2.2 Langevin Monte Carlo

Метод MALA (Metropolis Adjusted Langevin Algorithm), основанный на динамике Ланжевена, был предложен в [18, 29]. Уравнения Ланжевена исторически восходят к описанию броуновского движения малой частицы в вязкой среде. Первоначальный вид его был примерно таков:

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

где \mathbf{X} — положение частицы (а в нашем случае — трасса луча), t — время, U — потенциал, а $\boldsymbol{\xi}(t)$ выбирается независимо как нормально распределенный

случайный вектор с нулевым средним и единичной корреляционной матрицей (то есть компоненты вектора независимы). Стационарное распределение в таком процессе $\rho(\mathbf{X}) = Ze^{-U(\mathbf{X})}$, где Z — нормировка (стат. сумма). Очевидно, нам нужно, чтобы распределение было пропорционально функции значимости. Полного совпадения их быть не может, так как $\int \rho d\mathbf{X} = 1$, а функция значимости в общем случае не нормирована на 1.

Это уравнение можно интегрировать численно, например, с помощью явной разностной схемы Эйлера-Мариамы. Однако если «рельеф» потенциала сложный, содержит длинные узкие овраги и т.п., то для воспроизведения желаемого распределения с хорошей точностью требуется очень маленький шаг интегрирования по времени. То есть придется делать много шагов, и это долго. Ситуация усложняется, когда потенциал имеет разрывы, особенно такие, что на одной их стороне он бесконечен. Положение можно значительно улучшить, если использовать *обобщенное* уравнение Ланжевена [19, 30]:

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

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

где «сносовый член» (англ. drift term) $\mathbf{a}(\mathbf{X})$ имеет смысл «скатывающей силы», действующей на нашу частицу. Выражение $\sum_j \frac{\partial(\hat{T}\hat{T}^*)_{i,j}}{\partial X_j}$ является *дивергенцией* i -й строки $\hat{T}\hat{T}^*$, поэтому мы называем его *дивергентным членом*; \hat{T}^* — транспонированная матрица. Матрица $\hat{T}(\mathbf{X})$ осуществляет *локальное* преобразование координат (обычно с тем, чтобы избавиться от сильной анизотропии в исходном фазовом пространстве). Ее часто называют матрицей предварительной обработки (preconditioning matrix). Исторически она пришла из задач оптимизационного спуска [31, 32].

Идея ее построения проста: например, если пространство сжато по какой-то определённой координате, тогда, чтобы броуновский член $\hat{\xi}(t)$ позволял сделать удачное предложение перехода, не выходящее за область больших значений в функции вклада, по данной координате пространство необходимо «растянуть обратно». Ещё один пример сильной анизотропии — извилистый овраг потенциала. В такой ситуации матрица предварительной обработки должна «развернуть» пространство вдоль оврага и растянуть его «поперек оврага». Без такой матрицы почти любое предложение перехода будет пытаться «выскочить» из оврага и будет отвергнуто процедурой Гастингса.

Простейшим способом моделирования динамики Ланжевена является численное интегрирование по явной разностной схеме Эйлера-Мариамы, когда конечное смещение точки $\delta\mathbf{X}$ вычисляется по той же формуле, что и дифференциал:

$$\delta\mathbf{X} = \mathbf{a}(\mathbf{X})\delta t + \sqrt{2\delta t}\hat{T}\boldsymbol{\xi}(t), \quad (4)$$

где δt — *конечный* шаг по времени.

Найти новое положение точки $\hat{\mathbf{X}}$ при $t + \delta t$ по смещению $\delta\mathbf{X}$ непросто; как правило, это не будет $\hat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$. Проблема в том, что в большинстве случаев траектория ограничена определенным многообразием существенно меньшей размерности. Например, если трасса луча описывается через набор *направлений* своих N сегментов, то это многообразие — прямое произведение N поверхностей единичных сфер для этих направлений. Для мирового пространства путей узлы трассы луча лежат на поверхностях сцены и не могут сходить с них «в объем». Следовательно, в общем случае $\mathbf{X} + \delta\mathbf{X}$, лежащая в *касательной гиперплоскости*, «уходит» с разрешенного многообразия. Например, для направления сегмента луча $\mathbf{v} + \delta\mathbf{v}$ уже не будет *единичным* вектором. Поэтому вместо тривиального сложения $\hat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ необходимо еще и как-то вернуть точку $\mathbf{X} + \delta\mathbf{X}$ на «разрешенное» многообразие:

$$\hat{\mathbf{X}} = \mathcal{P}(\mathbf{X} + \delta\mathbf{X}). \quad (5)$$

Сделать это можно разными способами, например, просто взять *проекцию* точки на многообразие. Применительно к направлению сегмента луча — *нормировать вектор*, т.е. новое направление сегмента будет $\frac{\mathbf{v} + \delta\mathbf{v}}{|\mathbf{v} + \delta\mathbf{v}|}$.

Однако в случае, когда рельеф потенциала имеет «крутые склоны», такое численное интегрирование требует неприемлемо малого шага по времени. Причем при любом конечном δt стационарное распределение траектории не совпадает в *точности* с желаемым. Поэтому в методе MALA разностная схема дополняется правилом выбора Гастингса, преобразующим генерируемую *произвольной* марковской цепью последовательность так, чтобы она имела в *точности* желаемое распределение вероятностей. Метод позволяет работать даже с разрывным потенциалом с конечным, часто даже произвольно большим, шагом интегрирования по времени. Мы вычисляем смещение точки $\delta\mathbf{X}$ по (4), но вместо *безусловного* перехода в новое положение $\hat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ мы рассматриваем это лишь как *предложение* перехода, которое может быть принято (тогда траектория переходит в $\hat{\mathbf{X}}$) или нет (тогда траектория остается в точке \mathbf{X}). Вероятность принятия вычисляется как

$$\text{Pr} = \min \left(1, \frac{p(\mathbf{X}|\hat{\mathbf{X}})\rho(\hat{\mathbf{X}})}{p(\hat{\mathbf{X}}|\mathbf{X})\rho(\mathbf{X})} \right), \quad (6)$$

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. Это та же самая функция

с переставленными аргументами. При этом переход $\mathbf{X} \mapsto \hat{\mathbf{X}}$ является композицией двух отображений: (4) и (5). Вычисление вероятностей перехода для нашего случая дано в [7].

Необходимо понимать, что полученная в результате соединения разностной схемы и правила Гастингса марковская цепь — это совершенно другой объект, нежели исходное уравнение Ланжевена. Она имеет *в точности то же* стационарное распределение, но и только. Остальные статистические свойства (например, корреляционная функция, парные распределения и пр.) у нее могут быть сколь угодно отличны от исходного Ланжевена. Однако это не имеет почти никакого значения для применения в качестве генератора случайных точек в фазовом пространстве, т.к. нам от него нужно только желаемое распределение.

Для удешевления шага алгоритма МАLА предлагается несколько идей: (1) использовать диагональную матрицу предварительной обработки; (2) накапливать градиент и матрицу за несколько шагов алгоритма так, как это делает алгоритм оптимизации Adam [33]; (3) использовать “контролируемый” или адаптивный МСМС для правильного вычисления вероятностей перехода [34]. В работе [29] предлагаются ещё два метода: запоминание градиентов и гибридный метод накопления с запоминанием для получения более точной оценки усреднённых градиентов. Таким образом, предлагается, если можно так выразиться, некоторый “проинтегрированный градиент”, как единственный источник информации для формирования хорошего предложения перехода.

3 Мотивация и содержание нашего исследования

В настоящий момент работа [29] является терминальной точкой в применении гибридного Монте-Карло к задаче расчёта освещённости. На наш взгляд, эта работа делает шаг в верном направлении, т.к. применение алгоритмов на базе НМС возвращает нас к необходимости вычисления и хранения гессаана. Поэтому применение более продвинутых методов, таких как, например, RM-НМС [24], представляется нам интересной научной проблемой, но имеющей ряд фундаментальных трудностей для практического применения.

Алгоритм МАLА с матрицей предварительной обработки, напротив, перспективен, и при этом до конца не исследован в работах [18, 29]. А именно:

1. Не исследован вопрос работы с препятствиями или “потенциальными барьерами” в функции вклада. С точки зрения расчёта освещения — это теневое перекрытие одной поверхности другой. Градиент не может помочь эффективно обходить такие барьеры, т.к. просто не имеет информации о них. Эту информацию могла бы содержать корректно построенная матрица предварительной обработки.

2. В [18] уравнение Ланжевена записывается в отличном от оригинального уравнения виде, пренебрегая дивергентным членом. Влияние дивергентного члена на скорость сходимости MALA при различных условиях осталось неисследованным.
3. Диагональная форма матрицы предварительной обработки и способ её вычисления, предложенный в [18], являются интересным решением. Однако при увеличении размерности фазового пространства эффективность такого метода будет падать, т.к. невозможно сделать хорошее предложение перехода в произвольном направлении — диагональная матрица не может представить поворот.

Таким образом, при реализации алгоритма MALA есть три вычислительно ёмкие операции.

1. Вычисление матрицы предварительной обработки дорого, т.к. это $O(d^3)$ вычислений и $O(d^2)$ памяти, где d — размерность пространства. Однако на практике здесь возможны следующие возможности для оптимизации:
 - матрицу можно представить в блочно-диагональном виде;
 - элементы матрицы можно вычислять “лениво”, только по запросу, и при этом какие-то предварительные вычисления сделать заранее;
 - матрицу можно вычислять приближённо, точность на самом деле не очень важна для скорости сходимости метода.
2. Вычисление дивергентного члена является весьма дорогой операцией, а поскольку правило Гастингса обеспечивает в точности желаемое стационарное распределение для *произвольной* марковской цепи, то его, в принципе, можно и исключить из уравнения. Однако при этом может ухудшиться сходимость к стационарному распределению, например, снизится вероятность принятия перехода, так что генерируемая последовательность будет содержать много повторяющихся точек.
3. Аналогично можно попробовать исключить член с градиентом потенциала.

Мы исследуем влияние этих трёх членов (матрицы, дивергентного члена и градиентного члена) для того, чтобы выработать метод ускорения MALA в применении к задаче расчета освещённости.

Любая реализация MALA на практике столкнется с рядом неочевидных трудностей. Во-первых, необходимо каким-то образом выбрать фазовое пространство. От этого зависят почти все дальнейшие вычисления. Во-вторых, необходимо уметь выбирать величину шага по времени, что в случае MALA имеет ряд нюансов. В третьих, необходимо научиться вычислять плотности

вероятностей для правила Гастингса (6). Наконец, необходимо вычислять матрицу предварительной обработки. Это ключевые моменты, от которых зависит эффективность MALA. На них мы также подробно остановимся в нашем исследовании.

4 Влияние дивергентного члена

Сносный член включает в себя «дивергенцию» $\sum_j \frac{\partial(\hat{T}\hat{T}^*)_{i,j}}{\partial x_j}$, которая требует достаточно дорогостоящих вычислений. Для этого мало найти саму матрицу \hat{T} в данной точке, что уже достаточно непросто, так еще и требуется найти, как она от этой точки зависит.

Разумеется, в случае «настоящего» СДУ этот член принципиально важен и выбросить или заменить аппроксимацией его нельзя, так как в результате получится стационарное распределение. При этом сложно предсказать, насколько значительно. Поэтому необходимо вычисление этого выражения, причем максимально точное.

Единственным способом избавиться от этого члена было бы использовать постоянную матрицу преобразования \hat{T} , не зависящую от точки в фазовом пространстве. Однако это практически всегда приводит к другой трудности: теперь броуновская компонента не адаптирована под рельеф потенциала, и для интегрирования СДУ с достаточной точностью надо использовать совсем уж малый шаг по времени. Правда, при этом вычисления упрощаются, и каждый шаг считается быстрее, так что иногда (увы, редко) получается даже небольшое итоговое ускорение счета. Однако в большинстве случаев это приводит к фатальному замедлению.

К счастью, в случае, когда вместо «подлинной динамики СДУ» используется MALA, получающееся стационарное распределение будет точным безотносительно к тому, какие члены входят в СДУ. Можно выбросить даже градиент потенциала, не говоря уж о градиентном члене, сведя динамику к «Метрополису с броуновской мутацией»

$$d\mathbf{X} = \sqrt{2dt}\hat{T}\boldsymbol{\xi}(t), \quad (7)$$

и она все равно будет работать. Правда, скорость сходимости может зависеть от того, какие компоненты сносового члена оставлять. Однако в использованном нами для тестов примере даже и этого не произошло. Скорость сходимости не пострадала при отбрасывании дивергентного члена [8].

5 Выбор шага интегрирования

В численном интегрировании ОДУ шаг по времени является важнейшим параметром, и правильный, адаптивный (в зависимости от точки) его выбор определяет точность и скорость вычислений. В случае стохастических дифференциальных уравнений это даже еще более так. Правда, нам достаточно воспроизвести лишь стационарное распределение, а MALA обеспечивает *точное* воспроизведение этого распределения при *любом* шаге δt . Точнее, при *любом таком, что* траектория обходит всю область $\rho > 0$ за время вычислений. А при достаточно большом шаге это условие может *нарушаться*, как мы увидим ниже. Когда же это условие выполнено, то мы получим в точности желаемое распределение. Но остается вопрос о скорости сходимости к нему. Она же может сильно зависеть от шага по времени. В самом деле, при очень малом δt требуется очень большое число шагов $O(1/\delta t)$, так что вычисления будут очень медленными. С другой стороны, при большом δt обычно оказывается, что процедура Гастингса часто отвергает предложение перехода, *повторяя* текущее положение точки. Поэтому траектория оказывается сильно коррелированной (много зависимых значений), и это плохо при вычислении по ней средней для метода Монте-Карло.

Следовательно, даже и в MALA правильный выбор шага по времени важен. Какие неприятности происходят при неправильном выборе шага и как этого избежать, рассказано ниже.

5.1 Последствия слишком большого шага δt

Ниже мы пренебрегаем проекцией предложения (5) и полагаем, что новое положение точки есть просто $\hat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$. Для приближенных вычислений допустимого шага δt это позволительно.

При достаточно большом шаге по времени δt траектория может почти навечно застрять в какой-то точке. В самом деле, смещение точки под действием броуновского члена $\sqrt{2\delta t}\hat{T}\hat{\xi}$ не может соперничать с гораздо большим смещением под влиянием «скатывающей силы» $\mathbf{a}(\mathbf{X})\delta t$, поскольку оно растет гораздо быстрее с увеличением δt . Поэтому, хотя формально полное предложенное новое положение точки $\hat{\mathbf{X}} = \mathbf{X} + \delta\mathbf{X}$ и случайно, оно не может заметно отклониться от $\hat{\mathbf{X}}' \equiv \mathbf{X} + \mathbf{a}(\mathbf{X})\delta t$. Правило Гастингса (6) допускает переход в $\hat{\mathbf{X}}$, только если оттуда траектория *может вернуться назад*, т.е. если $\hat{\mathbf{X}} + \mathbf{a}(\hat{\mathbf{X}})\delta t' + \sqrt{2\delta t'}\hat{T}(\hat{\mathbf{X}})\boldsymbol{\eta} = \mathbf{X}$, где $\boldsymbol{\eta}$ — некий новый случайный вектор (вместо $\boldsymbol{\xi}$), а $\delta t'$ — временной шаг в точке $\hat{\mathbf{X}}$. Так как при большом $\delta t'$ снос доминирует над броуновским членом, то

$$\mathbf{X} \approx \hat{\mathbf{X}} + \mathbf{a}(\hat{\mathbf{X}})\delta t' \approx \mathbf{X} + \mathbf{a}(\mathbf{X})\delta t + \mathbf{a}(\hat{\mathbf{X}})\delta t',$$

что возможно, если только $\mathbf{a}(\hat{\mathbf{X}})$ не развернется в «обратную сторону», а это

в общем случае едва ли так. Поэтому *обратный переход невозможен* или его вероятность очень мала, и правило Гастингса запрещает *прямой* переход $\mathbf{X} \mapsto \hat{\mathbf{X}}$.

Эта ситуация изображена на левой половине Рисунка 1. А его правая половина изображает случай, когда, наоборот, шаг по времени мал и броуновское смещение может превысить снос.

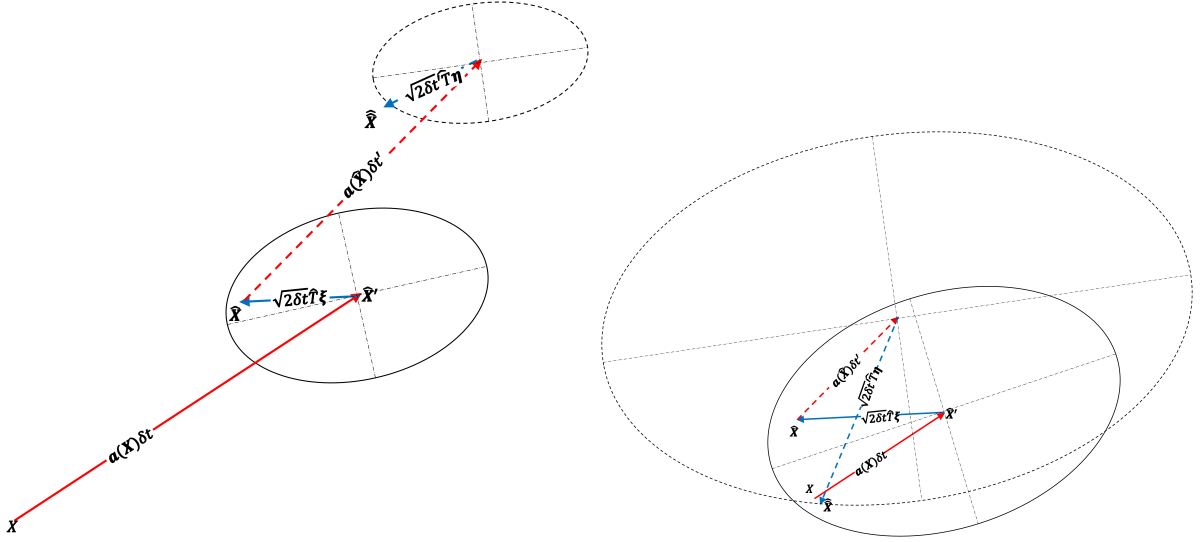


Рис. 1. На левой половине показан случай для достаточно большого δt (броуновский член мал по сравнению со сносом), а на правой — для маленького δt (броуновский член может превосходить снос). Красная стрелка показывает смещение за счет сносового члена, синяя стрелка — за счет броуновского члена, а эллипс ограничивает область, где после *обоих смещений* точка окажется с высокой вероятностью. Пунктирные стрелки и эллипс относятся к *следующему* шагу $\hat{\mathbf{X}} \mapsto \hat{\hat{\mathbf{X}}}$. Для показанного справа случая он *может вернуть* траекторию назад в \mathbf{X} , а слева — нет.

Следовательно, шаг по времени должен быть достаточно мал, чтобы броуновские флуктуации могли преодолеть смещение за счет скатывающей силы. Для этого *достаточно* (хотя обычно это избыточно сильное требование!), чтобы размер эллипса вдоль направления скатывающей силы был не меньше, чем длина стрелки, то есть $|\mathbf{a}(\mathbf{X})\delta t|$. Поэтому шаг по времени зависит от точки \mathbf{X} . Оказывается, что необходимо, чтобы выполнялось условие

$$\delta t \leq \min \left(2 \frac{|\hat{T}\mathbf{a}|^2}{|\mathbf{a}|^4}, \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right), \quad (8)$$

где ε — какое-то разумно малое число. Какое именно его значение брать, не столь уж важно, так как это все равно крайне приближенный вывод, в котором одна часть — избыточно достаточная, а вторая — минимально необходимая. Поэтому берем $\varepsilon = 0.1$ и трактуем (8) как *верхнюю границу*, при счете ис-

пользуем δt в 10 раз меньшее. Или берем минимум между тем, что дает (8), и каким-то эмпирическим ограничением наподобие $\delta t \leq 0.01$.

Из анализа влияния ограничений (раздел 6) формула расширяется до

$$\delta t \leq \min \left(\frac{1}{3}, \quad 2 \frac{|\hat{T}\mathbf{a}|^2}{|\mathbf{a}|^4}, \quad \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right). \quad (9)$$

Для надежности, конечно, лучше взять шаг поменьше, например, уменьшив даваемое вышеприведенной формулой значение раз в 10. Или добавить эмпирическое ограничение вычисленного по ней шага каким-то разумным пределом наподобие 0.01.

5.2 «Безопасный» шаг по времени (вывод формулы (8))

Как уже было сказано, шаг по времени должен быть достаточно мал, чтобы броуновские флуктуации могли преодолеть смещение за счет скатывающей силы. Для этого *достаточно*, чтобы размер эллипса вдоль направления скатывающей силы был не меньше, чем длина стрелки (рис. 1), то есть $|\mathbf{a}(\mathbf{X})\delta t|$.

Поскольку броуновское смещение $\sqrt{2\delta t}\hat{T}\boldsymbol{\xi}$ — это гауссовский случайный вектор с ковариационной матрицей $2\delta t\hat{T}^*\hat{T}$, то размер эллипса вдоль направления $\frac{\mathbf{a}}{|\mathbf{a}|}$ есть

$$3\sqrt{2\delta t \left(\frac{\mathbf{a}}{|\mathbf{a}|}, \hat{T}^*\hat{T} \frac{\mathbf{a}}{|\mathbf{a}|} \right)} = 3\sqrt{2\delta t} \frac{|\hat{T}\mathbf{a}|}{|\mathbf{a}|}.$$

Тройка возникает из-за «правила 3σ », то есть случайная величина почти никогда не превосходит трех дисперсий и соответствует размеру эллипса. И, по нашему требованию, эта величина должна быть не меньше, чем $|\mathbf{a}\delta t|$:

$$\sqrt{2\delta t} \frac{|\hat{T}\mathbf{a}|}{|\mathbf{a}|} \geq |\mathbf{a}\delta t|, \quad (10)$$

что и приводит к ограничению

$$\delta t \leq 2 \frac{|\hat{T}\mathbf{a}|^2}{|\mathbf{a}|^4}. \quad (11)$$

Это ограничение, хотя мы применяем его всегда, в действительности относится только к тем обычно нечастым случаям, когда предложенный переход может выбросить точку за пределы разрешенной области. Означает ли это, что выведенное выше ограничение избыточно, и можно двигаться со сколь угодно большим шагом, если разрешенная область очень велика? Формально в случае использования *полной* процедуры выбора Гастингса, обеспечивающей точное

совпадение стационарного распределения с желаемым, да, можно. Однако это может быть и неэффективно.

Очевидно, что если *исходная* выборка без применения процедуры Гастингса имеет стационарное распределение, *существенно* отличающееся от желаемого, то Гастингсу придется «приложить большие усилия» для выправления столь заметного несоответствия. Поэтому лучше ограничить шаг по времени так, что интегрирование уравнения Ланжевена без процедуры Гастингса давало стационарное распределение, не слишком отклоняющееся от желаемого.

Стационарное распределение ρ — это стационарное решение кинетического уравнения для марковской цепи, соответствующей разностной схеме интегрирования СДУ. Для достаточно гладкого распределения его эволюция в дискретном времени получается из-за, грубо говоря, свертки разности распределения в начальной и конечной точках шага, $\rho(\mathbf{X}(t + \delta t)) - \rho(\mathbf{X}(t))$ с распределением случайного члена. Когда эта разность может быть заменена *разложением распределения* в ряд Тэйлора до вторых производных включительно, тогда получается то кинетическое уравнение, которому удовлетворяет желаемое распределение. А если она не разлагается в ряд Тэйлора, то кинетическое уравнение будет другим, и желаемое распределение уже не будет его стационарным решением.

Очевидно, что для хорошей точности разложения в ряд Тэйлора необходимо хотя бы

$$|\rho(\mathbf{X} + \delta\mathbf{X}) - \rho(\mathbf{X})| \leq \varepsilon\rho(\mathbf{X})$$

для разумно малого ε , скажем, $\varepsilon = 0.1$. Оценивая левую часть с помощью градиента, приходим к критерию

$$|(\delta\mathbf{X}, \nabla \log \rho)| \leq \varepsilon,$$

или, поскольку стационарное распределение связано с потенциалом как $\rho(\mathbf{X}) = Ze^{-U(\mathbf{X})}$,

$$|(\delta\mathbf{X}, \nabla U)| \leq \varepsilon.$$

Если шаг по времени удовлетворяет (11), то смещение из-за скатывающей силы и из-за броуновского члена — одного порядка. Так что для приблизительного критерия можно взять $\delta\mathbf{X} \approx \mathbf{a}(\mathbf{X})\delta t$, что дает

$$|(\mathbf{a}, \nabla U)|\delta t \leq \varepsilon. \quad (12)$$

Соединяя этот критерий с (11), приходим к окончательному выражению (8).

6 Вычисление матрицы предварительной обработки из ограничений

Предложение перехода в любом из вариантов динамики (2), (3), (7) может переместить точку туда, где функция значимости обращается (или почти обращается

ется) в ноль, и такая мутация будет отвергнута процедурой Гастингса. Приемлемые переходы (т.е. те, которые могут быть приняты в процедуре Гастингса) формируют некую область *вокруг текущей точки*, т.е. трассы луча. Поэтому разумно устроить так, чтобы для почти всех предложенных переходов $\delta\mathbf{X}$ были бы автоматически ограничены этой областью. Иными словами, наложить *ограничения* на вектор $\delta\mathbf{X}$ или связи между его компонентами.

Наиболее типичные ограничения проистекают из, во-первых, наличия глянцевых поверхностей, чей BDF быстро исчезает при отклонении луча от направления зеркального отражения или преломления, и, во-вторых, ситуаций, когда луч рассеивается маленьким объектом сцены или проходит сквозь маленькое отверстие. В зависимости от сцены и трассы луча может действовать разное число ограничений.

Оказывается, что при малых смещениях все они могут быть сформулированы с помощью квадратичной формы вида

$$(\delta\mathbf{X}, \hat{Q}_k(\mathbf{X})\delta\mathbf{X}) \leq q_k^2, \quad k = 1, \dots, M,$$

и все эти неравенства должны выполняться одновременно. Конкретный вид матриц $\hat{Q}_k(\mathbf{X})$ зависит от сцены и от выбора представления трассы луча, т.е. от выбора фазового пространства. Их конкретный вид для представления трассы луча вычисляется через отклонения от зеркального направления [7].

Простейший способ объединить их в одно — это записать все соответствующие неравенства в форме $(\delta\mathbf{X}, q_k^{-2}\hat{Q}_k(\mathbf{X})\delta\mathbf{X}) \leq 1$ и просуммировать, что дает:

$$\begin{aligned} (\delta\mathbf{X}, \hat{Q}(\mathbf{X})\delta\mathbf{X}) &\leq 1 \\ \hat{Q}(\mathbf{X}) &\equiv \sum_{k=1}^M q_k^{-2}\hat{Q}_k(\mathbf{X}). \end{aligned} \tag{13}$$

Это излишне сильное ограничение. Например, если все \hat{Q}_k и q_k идентичны, итоговая матрица \hat{Q} будет завышена в M раз, требуя в \sqrt{M} меньших $\delta\mathbf{X}$ по сравнению с реально дозволенными. Поэтому такое грубое приближение имеет смысл лишь в том случае, когда необходимо ограничить $\delta\mathbf{X}$ в десятки и сотни раз по сравнению со свободным значением, когда без использования матрицы траектория вовсе не сдвинется с места.

В случае, когда шаг по времени удовлетворяет условию (11), полное смещение и броуновский член будут одного порядка величины, а потому достаточно ограничить не полное смещение $\delta\mathbf{X}$, но лишь его броуновскую компоненту. В этом случае подставляем в (13) $\delta\mathbf{X} = \sqrt{2\delta t}\hat{T}\boldsymbol{\xi}$ и приходим к

$$2\delta t(\boldsymbol{\xi}, \hat{T}^*\hat{Q}(\mathbf{X})\hat{T}\boldsymbol{\xi}) \leq 1. \tag{14}$$

Входящая в условие (14) \hat{Q} — неотрицательная симметричная матрица, для неё определен квадратный корень: $\hat{Q} = \left(\sqrt{\hat{Q}}\right)^* \sqrt{\hat{Q}}$, в терминах которого (14) принимает вид

$$(\boldsymbol{\xi}, \left(\sqrt{\hat{Q}\hat{T}}\right)^* \sqrt{\hat{Q}\hat{T}}\boldsymbol{\xi}) \leq \frac{1}{2\delta t},$$

что должно выполняться при практически всех $\boldsymbol{\xi}$.

Возможным решением будет $\hat{T} = c \left(\sqrt{\hat{Q}}\right)^{-1}$, и теперь $(\boldsymbol{\xi}, \boldsymbol{\xi})c^2 \leq \frac{1}{2\delta t}$. Поскольку $\boldsymbol{\xi}$ — нормально распределенный случайный вектор с единичной ковариацией, то большинство выборок удовлетворяют «правилу 3σ »: практически всегда $(\boldsymbol{\xi}, \boldsymbol{\xi}) \leq 3^2$, так что можно взять $c \leq \frac{1}{3\sqrt{2\delta t}}$, и все ограничения будут выполнены для большинства предложений перехода.

У нас теперь есть два условия, связывающие \hat{T} и δt : нынешнее и (8), и их необходимо объединить. Пусть коэффициент c есть константа, не зависящая от точки \mathbf{X} . Тогда, подставляя в (8) явное выражение для \mathbf{a}

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

и обозначая $\hat{\mathfrak{T}} \equiv \left(\sqrt{\hat{Q}}\right)^{-1}$, $\mathbf{a}_i = - \left(\hat{\mathfrak{T}}\hat{\mathfrak{T}}^*\nabla U\right)_i + \frac{\partial(\hat{\mathfrak{T}}\hat{\mathfrak{T}}^*)_{i,j}}{\partial X_j}$, так что $\mathbf{a} = c^2\mathbf{a}$, $\hat{T} = c\hat{\mathfrak{T}}$, приходим к

$$c^2\delta t \leq \min \left(2\frac{|\hat{\mathfrak{T}}\mathbf{a}|^2}{|\mathbf{a}|^4}, \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right).$$

С другой стороны, c должно удовлетворять условию $c \leq \frac{1}{3\sqrt{2\delta t}}$. Поэтому

$$c^2\delta t \leq \min \left(\frac{1}{18}, 2\frac{|\hat{\mathfrak{T}}\mathbf{a}|^2}{|\mathbf{a}|^4}, \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right).$$

Очевидно, можно взять любое c , и останется условие на δt вида (8). В наших расчетах исторически сложилось так, что $c = 1/\sqrt{6}$. Поэтому $\mathbf{a} = 6\mathbf{a}$, $\hat{\mathfrak{T}} = \sqrt{6}\hat{T}$, и

$$\hat{T} = \frac{1}{3\sqrt{2\delta t}} \left(\sqrt{\hat{Q}}\right)^{-1},$$

$$\delta t \leq c^{-2} \min \left(\frac{1}{18}, 2\frac{|\hat{\mathfrak{T}}\mathbf{a}|^2}{|\mathbf{a}|^4}, \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right) = \min \left(\frac{1}{3}, 2\frac{|\hat{T}\mathbf{a}|^2}{|\mathbf{a}|^4}, \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right).$$

Вычисление конкретного вида матриц квадратичных форм ограничений $\hat{Q}_k(\mathbf{X})$ зависит от сцены и от представления трассы луча. Таким образом, учет ограничений на трассу луча в сцене немного усиливает условие на шаг по времени (8), превращая его в (9)

$$\delta t \leq \min \left(\frac{1}{3}, \quad 2 \frac{|\hat{T}\mathbf{a}|^2}{|\mathbf{a}|^4}, \quad \frac{\varepsilon}{|(\mathbf{a}, \nabla U)|} \right).$$

7 Заключение

Стохастическая трассировка лучей с использованием механики Ланжевена выглядит весьма перспективным методом для расчета глобальной освещенности виртуальной сцены. Однако на сегодняшний день этот метод остается малоисследованным, в то время как его анализ может привести к значительному ускорению вычислений без потери качества получаемого результата.

Мы предприняли попытку проанализировать наиболее вычислительно емкие операции алгоритма MALA, постараться определить узкие места в нем и найти решения для их оптимизации. В первой части работы, представленной в данной статье, основное внимание было уделено оценке влияния дивергентного члена, выбору шага интегрирования и выводу формул для вычисления матрицы предварительной обработки, основанной на ограничениях. Показано, как эти аспекты алгоритма влияют на сходимость, т.е. на скорость его выполнения.

Дальнейшие работы [7, 8] будут направлены на анализ таких аспектов алгоритма MALA, как выбор фазового пространства, вычисление плотности вероятности предложения перехода и ограничений, налагаемых сценой на вариацию трассы, а также на проведение вычислительных экспериментов, подтверждающих результаты наших исследований.

Список литературы

- [1] Kajiya J.T. The rendering equation // Proceedings of the 13th annual conference on Computer graphics and interactive techniques (ACM SIGGRAPH 86). 1986. 20(4), pp. 143-150.
- [2] Фролов В.А, Волобой А.Г., Ершов С.В., Галактионов В.А. Современное состояние методов расчета глобальной освещенности в задачах реалистичной компьютерной графики // Труды Института системного программирования РАН. 2021. Т. 33. № 2. С. 7-48. [http://dx.doi.org/10.15514/ISPRAS-2021-33\(2\)-1](http://dx.doi.org/10.15514/ISPRAS-2021-33(2)-1)
- [3] Brooks S., Gelman A., Jones G., Meng X. Handbook of Markov Chain Monte Carlo. CRC press. 2011.

- [4] Sik M., Krivanek J. Survey of Markov Chain Monte Carlo Methods in Light Transport Simulation // IEEE Transactions on Visualization and Computer Graphics. 2018. Т. 26. № 4. pp. 1821-1840.
- [5] Veach E. Robust Monte Carlo methods for light transport simulation. — Ph.D. Thesis. Stanford University. 1998.
- [6] 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. <https://doi.org/10.20948/prepr-2023-63>
- [7] 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. <https://doi.org/10.20948/prepr-2023-64>
- [8] 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. <https://doi.org/10.20948/prepr-2023-65>
- [9] Veach E., Guibas L.J. Metropolis Light Transport // Proceedings of the 24th annual conference on Computer graphics and interactive techniques SIGGRAPH'97. 1997. pp. 65-76.
- [10] Kaplanyan A.S., Hanika J., Dachsbacher C. The Natural-constraint Representation of the Path Space for Efficient Light Transport Simulation. // ACM Transactions on Graphics (Proceedings of SIGGRAPH). 2014, Т. 33. № 4. Article 102. 13 pages.
- [11] Kelemen C., Szirmay-Kalos L., Antal G., Csonka F. A Simple and Robust Mutation Strategy for the Metropolis Light Transport Algorithm // Computer Graphics Forum (EUROGRAPHICS 2002). Oxford UK. 2002. Т. 21. № 3. pp. 531-540.
- [12] Hachisuka T., Kaplanyan A.S., Dachsbacher C. Multiplexed Metropolis Light Transport // ACM Transactions on Graphics (TOG), Proceedings of SIGGRAPH 2014. Т. 33. № 4. 10 pages.
- [13] Wenzel J. Light Transport on Path-Space Manifolds — Ph.D. Dissertation. Cornell University. 2013.
- [14] Hanika J., Droske M., Fascione L. Manifold next event estimation // Computer Graphics Forum. 2015, Т. 34. № 4. pp. 87-97.

- [15] Bitterli B., Wenzel J., Novak J., Jarosz W. Reversible Jump Metropolis Light Transport Using Inverse Mappings // ACM Transactions on Graphics (TOG). 2017. T. 37. № 1. 12 pages.
- [16] Otsu H., Kaplanyan A.S., Hanika J., Dachsbacher C., Hachisuka T. Fusing state spaces for markov chain Monte Carlo rendering // ACM Transactions on Graphics (TOG). 2017. T. 36. № 4. Article 74. 10 pages.
- [17] Pantaleoni J. Charted metropolis light transport. // ACM Transactions on Graphics (TOG). 2017. T. 36. № 4. Article 75. 14 pages.
- [18] Luan F., Zhao S., Bala K., Gkioulekas I. Langevin Monte Carlo rendering with gradient-based adaptation // ACM Transactions on Graphics (TOG). 2020. T. 39. № 4. Article 140. 16 pages.
- [19] Xifara T., Sherlock C., Livingstone S., Byrne S., Girolami M. Langevin diffusions and the Metropolis-adjusted Langevin algorithm // Statistics & Probability Letters. 2014. T. 91. C. 14-19. arXiv preprint arXiv:1309.2983.
- [20] Schreck A., Fort G., Moulines E. Adaptive Equi-nergy Sampler: Convergence and Illustration // ACM Transactions on Modeling and Computer Simulation (TOMACS). 2013. T. 23. № 1. Article 5. 27 pages. <https://doi.org/10.1145/2414416.2414421>
- [21] 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. T. 16. № 3. pp. 239-249. <https://doi.org/10.1007/s11222-006-8769-1>
- [22] Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo // arXiv preprint arXiv:1701.02434. — 2017.
- [23] Homan M.D., Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo // J. Mach. Learn. Res. 2014. T. 15. № 1. pp. 1593-1623.
- [24] Girolami M., Calderhead B., Chin S.A. Riemannian manifold hamiltonian monte carlo // arXiv preprint arXiv:0907.1100. — 2009.
- [25] Li T.-M., Lehtinen J., Ramamoorthi R., Wenzel J., Durand F. Anisotropic Gaussian mutations for metropolis light transport through Hessian-Hamiltonian dynamics // ACM Transactions on Graphics (TOG). 2015. T. 34. № 4. Article 209, 13 pages. <https://doi.org/10.1145/2816795.2818084>
- [26] Li T.-M. Differentiable Visual Computing. — Ph.D. Thesis. MIT. 2019.

- [27] Hogan R.J. Fast reverse-mode automatic differentiation using expression templates in C++ // ACM Transactions on Mathematical Software (TOMS). 2014. T. 40. № 4. Article 26. 16 pages. <https://doi.org/10.1145/2560359>
- [28] Rioux-Lavoie D., Litalien J., Gruson A., Hachisuka T., Nowrouzezahrai D. Delayed Rejection Metropolis Light Transport // ACM Transactions on Graphics (TOG). 2020. T. 39. № 3. Article 26, 14 pages. <https://doi.org/10.1145/3388538>
- [29] Luan F. Forward and inverse rendering with gradient based optimizations. — Ph.D. Thesis. Cornell University. 2021.
- [30] Zwanzig R. Nonlinear generalized Langevin equations // Journal of Statistical Physics. 1973. T. 9. № 3. pp. 215-220.
- [31] De Sa C. Lecture 8: Accelerating SGD with preconditioning and adaptive learning rates // CS4787 — Principles of Large-Scale Machine Learning Systems. Cornell University. <https://www.cs.cornell.edu/courses/cs4787/2019sp/notes/lecture8.pdf> (last accessed 23.11.2023)
- [32] Amari S., Ba J., Grosse R., Li X., Nitanda A., Suzuki T., Wu D., Xu J. When Does Preconditioning Help or Hurt Generalization? // arXiv preprint arXiv:2006.10732. — 2020.
- [33] Kingma D.P., Ba J. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. — 2014.
- [34] Andrieu C., Thoms J. A tutorial on adaptive MCMC // Statistics and computing. 2008. T. 18. pp. 343-373.