

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт (государственный университет)»
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Адаптивные методы семплирования по важности

Выпускная квалификационная работа
(бакалаврская работа)

Направление подготовки: 03.03.01 Прикладные математика и физика

Выполнил:

студент группы 474

Новицкий Василий Геннадьевич

(подпись обучающегося)

Научный руководитель:

Кандидат физико-математических наук

Максимов Юрий Владимирович

(подпись научного руководителя)

Москва 2018

Аннотация

В настоящее время методы Монте-Карло являются классическим инструментом в анализе надежности технических систем. К сожалению, их практическая эффективность при оценке редких событий крайне невысока. В данной работе нами предложены методы семплирования по важности для оценки вероятностей редких событий, а также приведены доказательства сходимости указанных методов. Численная эффективность методов проиллюстрирована на модельных данных.

Содержание

1	Введение	4
2	Постановка задачи	5
2.1	Модель постоянного тока	5
2.2	Приведение к стандартному многомерному нормальному распределению	6
2.3	Семплирование по важности	7
2.4	Смесь распределений	8
2.5	Оптимизационная постановка задачи	10
3	Метод Франка-Вульфа	12
4	Вычислительный эксперимент	16
5	Заключение	19
6	Список литературы	20

1 Введение

Международные цели по сокращению выбросов парниковых газов привели к буму в области возобновляемых источников энергии с особым акцентом на ветроэнергетику. В Европейском союзе поставлена цель — получить к 2020 году 20 процентов произведенной электроэнергии из возобновляемых источников энергии, причем часть энергии, полученной от ветровых источников, должна составить 35%.

Ветровые источники могут вносить флуктуации в электрическую сеть, что может привести к отказу сети. Поэтому возникает задача оценки надежности сети — найти вероятность отказа как можно более точно и быстро.

Чтобы описания поведения электрической сети можно использовать модель переменного тока. Модель дает нелинейную систему уравнений, которая описывает поток энергии по каждой линии передач. Из-за нелинейности анализ большой сети по модели переменного тока невозможен, и вместо этого используется линейная модель постоянного тока ([1]).

Энергопотребление случайных узлов электрической сети, таких как ветровой источник, будем моделировать с помощью нормального распределения. Эта аппроксимация адекватна для коротких по времени флуктуаций в сети.

Используя линеаризованную модель постоянного тока и нормальную модель энергопотребления случайных узлов, мы сводим задачу нахождения вероятности отказа к следующей: найти вероятность вылета нормального распределенного случайного вектора за полиэдр. Задача решается численно методами семплирования по важности, потому что дисперсия оценки, полученной методами семплирования по важности, меньше дисперсии оценки, полученной методами Монте-Карло. Был проведен вычислительный эксперимент на модельных данных, в котором показана численная эффективность метода.

Структура работы выстроена следующим образом. В разделе 2 формулируется оптимизационная постановка задачи. Методы, используемые для решения задачи оптимизации, приведены в разделе 3. Результаты вычислительного эксперимента приводятся в разделе 4. Раздел 5 — заключительный, в нем резюмируются основные выводы.

2 Постановка задачи

2.1 Модель постоянного тока

Описание сети приведено, как в [2], [3]. Наша электрическая сеть состоит из N узлов. Узел может как производить энергию, так и потреблять ее. Также имеется M линий передач. Обычно сеть разреженная, а M – в константу раз больше, чем N .

Пусть p_i – производство электроэнергии на i -том узле. Будем считать, что у потребителей $p_i < 0$. У некоторых узлов производство энергии p_i задается однозначно, мы будем называть их постоянными (fixed) узлами. У некоторых узлов производство энергии случайное, мы будем называть их случайными (random) узлами. Также имеется балансирующий (slack) узел: у него $p_S = \sum_{i \neq S} p_i$, необходимый для выполнения уравнения баланса мощностей:

$$\sum_i p_i = 0. \quad (1)$$

Производство энергии на каждом из узлов можно объединить в вектор $p = (p_F^\top, p_R^\top, p_S)^\top$, где p_F – вектор, соответствующий постоянным (fixed) узлам, p_R – вектор, соответствующий случайным (random) узлам, p_S – производство балансирующего (slack) узла. Пусть у нас имеется N_F постоянных узлов, N_R случайных узлов и $N_S = 1$ балансирующий узел. Тогда $N = N_F + N_R + N_S$.

Энергопотребление случайных узлов будем моделировать нормальным распределением: $p_R \sim \mathcal{N}(\eta_R, \Sigma_R)$. Мощность каждого случайного узла должна удовлетворять ограничениям:

$$\underline{p}_i^R \leq p_i^R \leq \overline{p}_i^R. \quad (2)$$

Энергопотребление постоянных узлов удовлетворяет равенству: $p_F = \eta_F$.

Мощность балансирующего узла должна удовлетворять ограничениям:

$$\underline{p}_S \leq p_S \leq \overline{p}_S.$$

Из уравнения баланса 1 следует, что $p_S \sim \mathcal{N}(\eta_S, \Sigma_S)$, где $\eta_S = -1_R^\top \eta_R - 1_F^\top \eta_F$. Поэтому ограничение на p_S можно переписать в виде ограничений на p_R :

$$\underline{p}_S + 1_F^\top p_F \leq -1_R^\top p_R \leq \overline{p}_S + 1_F^\top p_F. \quad (3)$$

Фазу на узле i обозначим θ_i , θ – вектор фаз на узлах. В модели постоянного тока приближенно выполняется равенство: $B\theta = p$, где матрица B (лапласиан) известна. Зная вектор мощностей p , можем найти $\theta = B^+p$. Фазовые ограничения на сеть:

$$|\theta_i - \theta_j| \leq \overline{\theta_{ij}}, \quad i \neq j. \quad (4)$$

Во всех наших примерах $\overline{\theta_{ij}} = \bar{\theta}$.

Выражая из уравнения баланса (1) $p_S = -1_R^T p_R - 1_F^T p_F$, учитывая $p_F = \eta_F$ и подставляя это в (4), получаем $2M$ (по 2 на линию в сети) линейных ограничений на вектор p_R .

Всего получается $2N_R$ ограничений от случайных узлов (2), 2 ограничения от балансирующего узла (3), $2M$ фазовых ограничений (4) на вектор p_R , распределенный нормально.

Задачу нахождения вероятности сбоя электрической сети можно переформулировать так: найти вероятность того, чтобы хотя бы одно из $J = 2M + 2N_R + 2$ ограничений на вектор $p_R \sim \mathcal{N}(\eta_R, \Sigma_R)$ нарушится. Переобозначим $x \equiv p_R$, $\eta \equiv \eta_R$, $\Sigma \equiv \Sigma_R$. Пусть H_j – событие, заключающееся в том, что j -ое из J ограничений нарушено. Тогда искомая вероятность

$$\mu = \mathbb{P} \left(x \in \bigcup_{j=1}^J H_j \right),$$

где $H_j = \{x \mid w_j^T x \geq \tau_j\}$ – линейное ограничение.

2.2 Приведение к стандартному многомерному нормальному распределению

Пусть имеется нормально распределенный вектор $x \in \mathbb{R}^d$, $x \sim \mathcal{N}(\eta, \Sigma)$ и линейные ограничения $H_j = \{x \mid w_j^T x \geq \tau_j\}$. Необходимо посчитать

$$\mu = \mathbb{P} \left(x \in \bigcup_{j=1}^J H_j \right).$$

Заметим, что линейная замена $x \leftarrow \Sigma^{-1/2}(x - \eta)$ приводит к тому, что $x \sim \mathcal{N}(0, I_d)$. Ограничения H_j при этом изменяются следующим образом:

$$\tau_j \leftarrow \frac{\tau_j - w_j^T \eta}{\sqrt{w_j^T \Sigma w_j}},$$

$$w_j \leftarrow \frac{\Sigma^{1/2} w_j}{\sqrt{w_j^T \Sigma w_j}}$$

и остаются линейными.

Задача свелась к нахождению вероятности

$$\mu = \mathbb{P} \left(x \in \bigcup_{j=1}^J H_j \right), \quad (5)$$

где $x \sim \mathcal{N}(0, I_d)$. Линейные ограничения – $H_j = \{x \mid w_j^T x \geq \tau_j\}$. Пересечение дополнений этих ограничений образует полиэдр \mathcal{P} . Нахождение (5) – это нахождение вероятности вылета стандартного многомерного нормально распределенного вектора x за полиэдр \mathcal{P} .

2.3 Семплирование по важности

Пусть $H = \bigcup_{j=1}^J H_j$. Перепишем задачу (5) в виде:

$$\mu = \mathbb{P}(x \in H) = \int f(x) p(x) dx,$$

где $f(x) = \mathbb{I}(x \in H)$ – индикаторная функция множества H , $p(x)$ – плотность распределения нормального вектора x .

Обычно такие интегралы оцениваются с помощью метода Монте-Карло (см. [4]):

$$\hat{\mu}^{MC} = \frac{1}{n} \sum_{i=1}^n f(x_i), \quad x_i \sim p(x)$$

Дисперсия и коэффициент вариации оценки Монте-Карло:

$$\text{Var } \hat{\mu}^{MC} = \frac{\sigma^2}{n}, \quad \text{где } \sigma = \sqrt{\mu(1-\mu)},$$

$$\text{cv} = \frac{\sigma/\sqrt{n}}{\mu} = \frac{\sqrt{\mu(1-\mu)}}{\mu\sqrt{n}} \sim \frac{1}{\sqrt{\mu n}}.$$

Событие «отказ сети» – редкое (вероятность μ порядка $10^{-5} - 10^{-7}$), поэтому для получения коэффициента вариации хотя бы 0.1, необходимо сгенерировать $n = 100\mu^{-1} \sim 10^7 - 10^9$ точек, что очень много.

Идея семплирования по важности состоит в том, что некоторые значения случайного вектора генерировать более важно, чем другие. Оказывается, что

для снижения дисперсии в данной задаче нужно чаще генерировать точки вне полиэдра \mathcal{P} .

Будем независимо генерировать точки x_i , $i = 1, \dots, n$ не из исходного распределения $p(x)$, как в методе Монте-Карло, а из смещенного распределения $q(x)$. Распределение $q(x)$ должно удовлетворять условиям: $q(x) > 0$ везде, где $f(x)p(x) \neq 0$. Оценка метода семплирования по важности:

$$\widehat{\mu}_q^{IS} = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)p(x_i)}{q(x_i)}, \quad x_i \sim q. \quad (6)$$

Легко проверить, что оценка (6) несмещенная, то есть $\mathbf{E}\widehat{\mu}_q^{IS} = \mu$. Дисперсия оценки:

$$\text{Var}(\widehat{\mu}_q^{IS}) = \frac{1}{n} \left(\int \frac{(fp)^2}{q} dx - \mu^2 \right) = \frac{1}{n} \int \frac{(fp - \mu q)^2}{q} dx \quad (7)$$

Чем меньше дисперсию мы получим, тем меньшим количеством семплов мы сможем обойтись для оценки вероятности отказа с заданной точностью (дисперсией оценки). Для минимизации дисперсии предлагается ввести параметризацию распределения q и решить задачу минимизации дисперсии численными методами оптимизации.

2.4 Смесь распределений

Мы используем технику представления q в виде смеси распределений (Mixture Importance Sampling, MIS) из [5] и [6]:

$$q_\alpha(x) = \sum_{j=1}^J \alpha_j q_j(x). \quad (8)$$

Коэффициенты α_j имеют смысл вероятностей, с которыми выбирается одно из распределений q_j , из которого будет сгенерирована точка, поэтому $\sum_{j=1}^J \alpha_j = 1$.

В [2] был предложен метод ALOE. В этом методе в качестве базисных распределений q_j берутся условные распределения $q_j = \mathcal{L}(x | H_j)$. Так как вероятностные меры q_j отличны от нуля только за пределами полиэдра \mathcal{P} , то все точки из распределения q генерируются только вне этого полиэдра. Возможность эффективно генерировать точки из условных распределений существенно

опирается на свойства многомерного нормального распределения. О том, как семплировать точки из условного нормального распределения, написано в [2]. Введем условные распределения. Плотность распределения q_j :

$$q_j(x) = \frac{H_j(x)p(x)}{P_j}, \quad (9)$$

где $H_j(x) = \mathbb{I}(x \in H_j)$, $P_j = \mathbf{P}(H_j)$.

В методе ALOE предлагается взять коэффициенты

$$\alpha_j^{ALOE} = P_j \left(\sum_{k=1}^J P_k \right)^{-1}.$$

Такии образом, получается оценка ALOE:

$$\hat{\mu}^{ALOE} = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)p(x_i)}{q_\alpha(x_i)} = \frac{\sum_{j=1}^J P_j}{n} \sum_{i=1}^n \frac{\mathbb{I}(x_i \in H)}{\sum_{j=0}^J H_j(x_i)} \quad (10)$$

Пусть $S(x) = \sum_j \mathbb{I} x \in H_j$. В [2] доказана теорема:

$$\text{Var}(\hat{\mu}^{ALOE}) \leq \frac{\mu^2 J + J^{-1} - 2}{n} \frac{1}{4}$$

Однако эта оценка точная, если распределение $S(x)$ равномерное на $1, \dots, J$. Можно придумать полиэдр, когда это не так. Пусть наш полиэдр \mathcal{P} – это правильная пирамида с маленьким углом при основании и большим числом граней J . Пусть вероятности вылететь за каждое ограничение одна и та же. Тогда $H_j = \{x \mid w_j^\top x \geq \tau_j\}$, где

$$w_j = \begin{cases} (0, 0, -1), & \text{если } j = 1, \\ \left(\cos\left(\frac{2\pi(j-1)}{J-1}\right) \sin \theta, \sin\left(\frac{2\pi(j-1)}{J-1}\right) \sin \theta, \cos \theta \right), & \text{если } j = 2, \dots, J; \end{cases}$$

$$\tau_j = \tau, \quad j = 1, \dots, J.$$

В этом случае коэффициенты α_j^{ALOE} все одинаковые, так как все P_j одинаковые. Если угол при основании пирамиды мал, то каждый раз, когда мы генерируем точку из q_j , где $j \geq 2$, то она с вероятностью, близкой к 1, вылетает за все боковые грани сразу.

Пусть $U = \mathbb{I}\left(\left(x \in \bigcup_{j=2}^J H_j\right) \cap x \notin H_1\right)$. В силу малости угла θ $\mathbf{P}(x \in U) \approx \mathbf{P}(x \in H_1)$. Однако получается, что точки из U генерируются примерно в $J-1$ раз чаще, чем точки из H_1 , а интуиция говорит, что надо поровну.

Эти соображения приводят нас к тому, что нужно подбирать коэффициенты α_j , учитывая геометрию полиэдра.

2.5 Оптимизационная постановка задачи

Подставим в формулу для дисперсии оценки методом семплирования по важности (7) распределение $q(x)$ как смесь (8) условных (9) распределений. При этом в отличие от ALOE, не будем уточнять коэффициенты α_j и ограничимся их нормировкой на единицу:

$$\text{Var}(\widehat{\mu}_\alpha) = \frac{1}{n} \left(\int \frac{H_{1..J}(x)}{\sum_{j=1}^J \alpha_j H_j(x) P_j^{-1}} dx - \mu^2 \right) \rightarrow \min, \quad (11)$$

$$\sum_{j=1}^J \alpha_j = 1, \quad \alpha_j \geq \varepsilon, \quad \varepsilon \approx 0.1/J. \quad (12)$$

Здесь ограничения $\alpha_j \geq \varepsilon$ введены, как в [5], для того чтобы знаменатель в (11) не обращался в ноль. Мы берем $\varepsilon \approx 0.1/J$, как в [5].

Будем минимизировать не сам функционал (дисперсию), а его несмещенную оценку (эмпирическую дисперсию) (это обычно бывает эффективнее, см. [7]):

$$\widehat{\text{Var}}(\widehat{\mu}_\alpha) = \frac{1}{n} \left(\sum_{i=1}^n \frac{H_{1..J}(x_i)}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1} \right) \left(\sum_{j=1}^J \alpha'_j H_j(x_i) P_j^{-1} \right)} - \mu^2 \right) \rightarrow \min,$$

$$x_i \sim q_{\alpha'}, \quad q_{\alpha'}(x) = \sum_{j=1}^J \alpha'_j H_j(x) P_j^{-1} p(x),$$

где α' – некоторые заранее выбранные коэффициенты, удовлетворяющие (12). Минимизация эмпирической дисперсии эквивалентна минимизации следующего функционала :

$$f(\alpha) = \widehat{\text{Var}}(\widehat{\mu}_\alpha) + \frac{\mu^2}{n}.$$

Окончательная постановка оптимизационной задачи:

$$f(\alpha) = \frac{1}{n} \sum_{i=1}^n \frac{H_{1..J}(x_i)}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1} \right) \left(\sum_{j=1}^J \alpha'_j H_j(x_i) P_j^{-1} \right)} \rightarrow \min,$$

$$\sum_{j=1}^J \alpha_j = 1, \alpha_j \geq \varepsilon, \quad (13)$$

$$x_i \sim q_{\alpha'}, \quad q_{\alpha'}(x) = \sum_{j=1}^J \alpha'_j H_j(x) P_j^{-1} p(x).$$

Введем обозначения:

$$f_i(\alpha) = \frac{H_{1:J}(x_i)}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1}\right) \left(\sum_{j=1}^J \alpha'_j H_j(x_i) P_j^{-1}\right)}.$$

Тогда $f(\alpha)$ из (13) можно записать в виде :

$$f(\alpha) = \frac{1}{n} \sum_{i=1}^n f_i(\alpha).$$

Теорема 1 Пусть $S = \left\{ \alpha \mid \sum_{j=1}^J \alpha_j = 1, \alpha_j \geq \varepsilon > 0 \right\}$. Тогда функция $f(\alpha)$ – выпуклая на множестве S .

Доказательство:

$$f(\alpha) = \frac{1}{n} \sum_{i=1}^n f_i(\alpha) = \sum_{i=1}^n c_i h_i(\alpha), \text{ где}$$

$$h_i(\alpha) = \frac{1}{\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1}},$$

$$\frac{\partial^2 h_i(\alpha)}{\partial \alpha_p \partial \alpha_q} = \frac{2 H_p(x_i) P_p^{-1} H_q(x_i) P_q^{-1}}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1}\right)^3}.$$

Введем вектор $v_i = \left(v_i^{(1)}, \dots, v_i^{(J)} \right)^\top$, где $v_i^{(j)} = H_j(x_i) P_j^{-1}$. Тогда

$$\nabla^2 h_i(\alpha) = \frac{2}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1}\right)^3} v_i v_i^\top \succcurlyeq 0.$$

Функция $h_i(\alpha)$ – выпуклая, так как ее гессиан неотрицательно определен. Тогда $f(\alpha)$ – выпуклая как линейная комбинация выпуклых функций с неотрицательными коэффициентами.

3 Метод Франка-Вульфа

Для решения оптимизационной задачи (13) предлагается применить алгоритм Франка-Вульфа, который эффективен при минимизации выпуклых функций на полиэдре. В нашей оптимизационной задаче функция $f(\alpha)$ выпуклая на множестве $S = \left\{ \alpha \mid \sum_{j=1}^J \alpha_j = 1, \alpha_j \geq \varepsilon > 0 \right\}$, и это множество является полиэдром. Вариант метода Франка-Вульфа представлен в алгоритме 3.1.

Алгоритм 3.1 Алгоритм Франка-Вульфа

Вход: начальное значение аргумента $\alpha^{(0)}$, T ;

- 1: **Цикл** $t = 0, \dots, T - 1$ **выполнять**
 - 2: $v_t = \arg \max_{v \in S} \langle v, -\nabla f(\alpha^{(t)}) \rangle$
 - 3: $d_t = v_t - \alpha^{(t)}$
 - 4: $\gamma_t = \frac{2}{t+1}$
 - 5: $\alpha^{(t+1)} = \alpha^{(t)} + \gamma_t d_t$
 - 6: **Конец цикла**
 - 7: **Возвратить** $\alpha^{(T)}$
-

Теорема 2 Функция f из (13) – липшицева на $S = \left\{ \alpha \mid \sum_{j=1}^J \alpha_j = 1, \alpha_j \geq \varepsilon > 0 \right\}$ с константой $L = \frac{2P_{\max}^4}{\varepsilon^4} \sum_{j=1}^J P_j^{-2}$.

Доказательство:

Используем обозначения из теоремы 1.

$$f(\alpha) = \frac{1}{n} \sum_{i=1}^n f_i(\alpha).$$

Пусть L_i – константа Липшица функции h_i . Введем вектор $v_i = \left(v_i^{(1)}, \dots, v_i^{(J)} \right)^\top$, где $v_i^{(j)} = H_j(x_i) P_j^{-1}$. Тогда

$$\nabla^2 h_i(\alpha) = \frac{2}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1} \right)^3} v_i v_i^\top \succcurlyeq 0.$$

Оценим L_i^h сверху максимальным собственным значением гессиана:

$$L_i^h \leq \lambda_{\max} (\nabla^2 h_i(\alpha)) = \frac{2}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1}\right)^3} \|v_i\|^2. \quad (14)$$

Оценим знаменатель:

$$\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1} \geq \varepsilon \sum_{j=1}^J H_j(x_i) P_j^{-1} \geq \varepsilon P_{\max}^{-1}. \quad (15)$$

Оценим норму v_i :

$$\|v_i\|^2 = \sum_{j=1}^J H_j(x_i) P_j^{-2} \leq \sum_{j=1}^J P_j^{-2}. \quad (16)$$

Подставляя (15) и (16) в (14), получим:

$$L_i^h \leq \lambda_{\max} (\nabla^2 h_i(\alpha)) = \frac{2P_{\max}^3}{\varepsilon^3} \sum_{j=1}^J P_j^{-2}.$$

Теперь получим оценку на константу липшица функции $f_i(\alpha)$:

$$f_i(\alpha) = \frac{H_{1:J}(x_i)}{\left(\sum_{j=1}^J \alpha_j H_j(x_i) P_j^{-1}\right) \left(\sum_{j=1}^J \alpha'_j H_j(x_i) P_j^{-1}\right)}$$

$$L_i^f = \frac{H_{1:J}(x_i)}{\left(\sum_{j=1}^J \alpha'_j H_j(x_i) P_j^{-1}\right)} L_i^h \leq L_i^h \frac{P_{\max}}{\varepsilon} = \frac{2P_{\max}^4}{\varepsilon^4} \sum_{j=1}^J P_j^{-2}.$$

Теорема 3 Пусть функция f – выпуклая и липшицева с константой L . Тогда $f(\alpha^{(k)}) \leq \frac{4L}{k+2}$.

Доказательство приведено в [8]

На основе методов SVRF и SAGA в [9], [10] были предложены стохастические варианты метода Франка-Вульфа, их схемы приведены в алгоритмах 3.2 и 3.3.

Алгоритм 3.2 Стохастический метод Франка-Вульфа – SVRF-FW

Вход: начальное значение аргумента $\alpha_m^{(0)} = \alpha^{(0)}$, количество итераций T , размер эпохи m , количество эпох $S = \lfloor T/m \rfloor$, стратегия выбора шага $\{\gamma_i\}_{i=0}^{m-1}$, где $\gamma_i \in [0, 1]$, размеры мини-батчей $\{b_i\}_{i=0}^{m-1}$;

1: **Цикл** $s = 0, \dots, S - 1$ **выполнять**

2: $\tilde{\alpha}_s = \alpha_s^{(m)}$

3: $\tilde{g}_s = \nabla f(\tilde{\alpha}_s)$

4: **Цикл** $t = 0, \dots, m - 1$ **выполнять**

5: Выбрать подмножество I_t размера b_t равномерно с возвращениями из $\{1, \dots, n\}$

6: $g_{s+1}^{(t)} = \frac{1}{b_t} \sum_{i \in I_t} \left(\nabla f_i(\alpha_{s+1}^{(t)} - \nabla f_i(\tilde{\alpha}_s) + \tilde{g}_s) \right)$

7: $v_{s+1}^{(t)} = \arg \max_{v \in S} \langle v, -g_{s+1}^{(t)} \rangle$

8: $d_{s+1}^{(t)} = v_{s+1}^{(t)} - \alpha_{s+1}^{(t)}$

9: $\alpha_{s+1}^{(t+1)} = \alpha_{s+1}^{(t)} + \gamma_t d_{s+1}^{(t)}$

10: **Конец цикла**

11: **Конец цикла**

12: **Возвратить** точка, выбранная равномерно из $\left\{ \left\{ \alpha_{s+1}^{(t)} \right\}_{t=0}^{m-1} \right\}_{s=0}^{S-1}$

Алгоритм 3.3 Стохастический метод Франка-Вульфа – SAGA-FW

Вход: начальное значение аргумента $\beta_i^{(0)} = \alpha^{(0)}, \forall i \in \{1, \dots, n\}$, количество итераций T , стратегия выбора шага $\{\gamma_i\}_{i=0}^{T-1}$, где $\gamma_i \in [0, 1]$, размеры мини-батчей $\{b_i\}_{i=0}^{T-1}$;

Выход: $\alpha^{(T)}$.

1: $g^{(0)} = \frac{1}{n} \sum_{i=1}^n \nabla f(\beta_i^{(0)})$

2: **Цикл** $t = 0, \dots, T - 1$ **выполнять**

3: Выбрать подмножества I_t и J_t размера b_t равномерно с возвращениями из $\{1, \dots, n\}$

4: $g^{(t)} = \frac{1}{b_t} \sum_{i \in I_t} (\nabla f_i(\alpha^{(t)} - \nabla f_i(\beta_i^{(t)}) + \tilde{g}^{(t)})$

5: $v^{(t)} = \arg \max_{v \in S} \langle v, -g^{(t)} \rangle$

6: $d^{(t)} = v^{(t)} - \alpha^{(t)}$

7: $\alpha^{(t+1)} = \alpha^{(t)} + \gamma_t d^{(t)}$

8: $\beta_j^{(t+1)} = \alpha^{(t)}$, если $j \in J_t$ или $\beta_j^{(t+1)} = \beta_j^{(t)}$, если $j \notin J_t$

9: $g^{(t+1)} = g^{(t)} - \frac{1}{n} \sum_{j \in J_t} (\nabla f_j(\beta_j^{(t)}) - \nabla f_j(\beta_j^{(t+1)}))$

10: **Конец цикла**

11: **Возвратить** точка, выбранная равномерно из $\{\alpha^{(t)}\}_{t=0}^{T-1}$

К сожалению, сходимость методов доказана только для шага, обратно пропорционального константе Липшица L , что приводит к нулевому шагу.

4 Вычислительный эксперимент

Эксперимент был проведен на модельных данных (пирамиде) из раздела 2.4: $H_j = \{x \mid w_j^\top x \geq \tau_j\}$, где

$$w_j = \begin{cases} (0, 0, -1), & \text{если } j = 1, \\ \left(\cos\left(\frac{2\pi(j-1)}{J-1}\right) \sin \theta, \sin\left(\frac{2\pi(j-1)}{J-1}\right) \sin \theta, \cos \theta \right), & \text{если } j = 2, \dots, J; \end{cases}$$

$$\tau_j = \tau, \quad j = 1, \dots, J.$$

Параметры пирамиды: число граней $J = 1000$, угол при основании $\theta = 0.01$, параметр, определяющий размер пирамиды $\tau = 4$. Между собой сравнивались методы Франка-Вульфа, стохастические методы Франка-Вульфа: SVRF-FW и SAGA-FW, а так же метод проекции градиента (на графиках обозначен GD).

Были построены графики сходимости по норме аргумента, по главной компоненте аргумента и по функции от времени. Под нормой аргумента подразумевается L2-норма разности текущего аргумента и оптимального. Оптимальный аргумент находится одномерным тернарным поиском (одномерным, потому что силу правильности пирамиды $\alpha_1 \equiv \alpha$, $\alpha_2^* = \alpha_3^* = \dots = \alpha_J^* = (1 - \alpha) / (J - 1)$, тернарным, потому что оптимизируемый функционал (13) выпуклый).

Под главной компонентой аргумента подразумевается α_1 . В силу соображений, упомянутых в разделе 2.4, понятно, что $\alpha_1^* \approx 0.5$. Точное значение было найдено тернарным поиском.

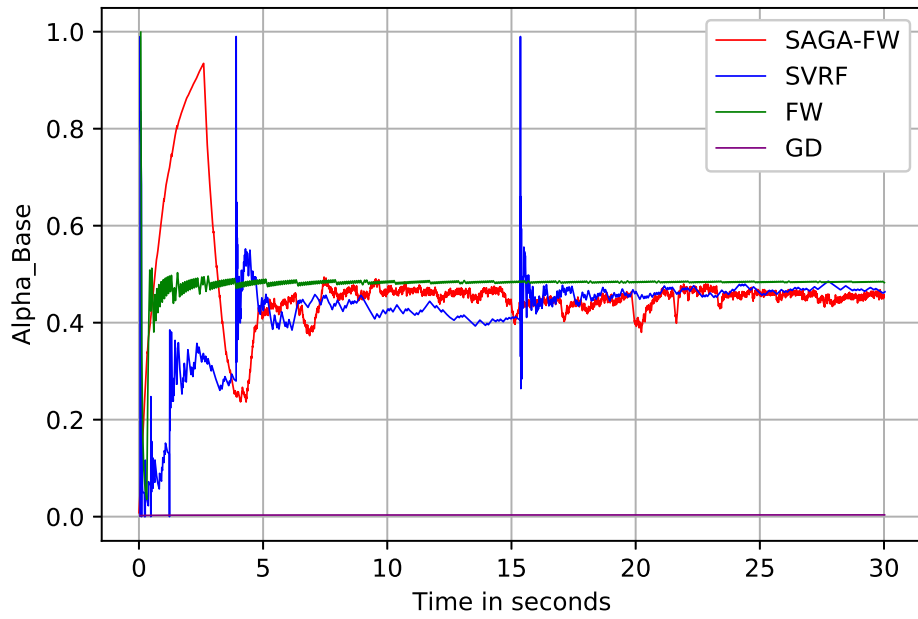


Рис. 1: График сходимости по главной компоненте аргумента от времени

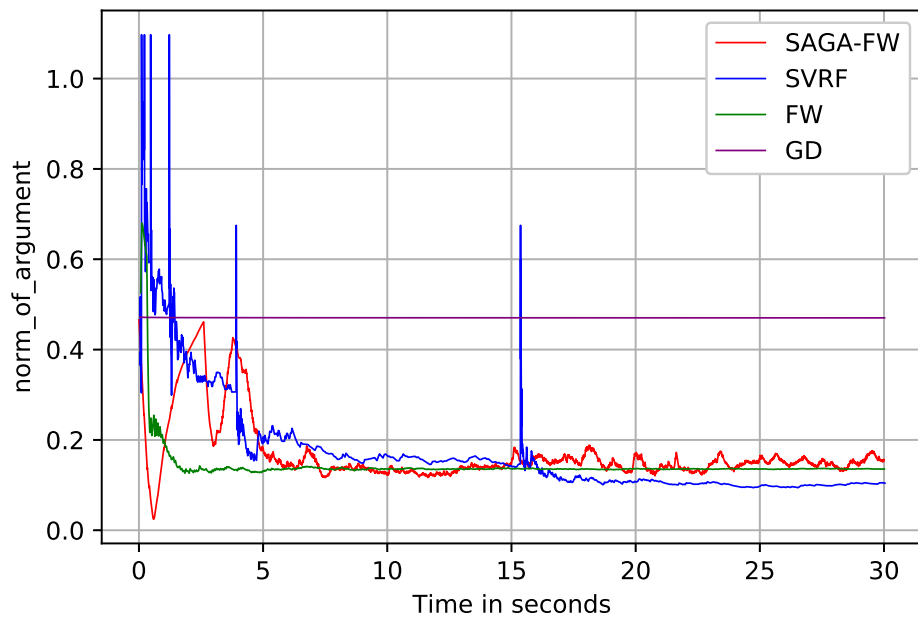


Рис. 2: График сходимости по норме отклонения аргумента от оптимума от времени

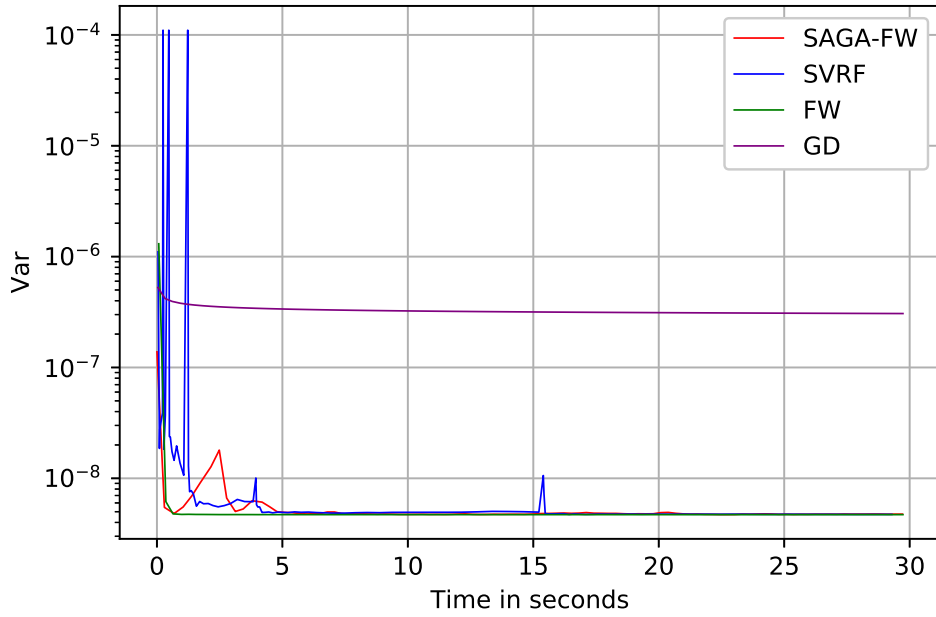


Рис. 3: График сходимости по функции от времени

Метод	SAGA-FW	SVRF	FW	GD
α_1	0.52	0.48	0.51	0.04
Var	5.08×10^{-9}	5.11×10^{-9}	5.06×10^{-9}	3.59×10^{-7}
cv	1.2	1.2	1.2	9.9

Таблица 1: Результаты после 30 секунд прогона

Из графиков видно, что методы Франка-Вульфа сходятся гораздо быстрее, чем метод проекции градиента, так как они более полно учитывают структуру ограничений (полиэдр) в задаче. Из таблицы видно, что за время эксперимента градиентный спуск практически не уменьшил дисперсию оценки, тогда как методы Франка-Вульфа улучшили дисперсию в 100 раз. Видно, что стохастические варианты метода Франка-Вульфа не дают выигрыша по времени, по сравнению с обычным методом Франка-Вульфа, это связано с трудностью настройки шага в них, так как теоремы, доказывающие сходимость используют шаг, пропорциональный константе Липшица, а у нас есть только сильно завышенная её оценка.

5 Заключение

В работе были предложены алгоритмы, улучшающие методы Монте-Карло для оценки вероятности редких событий. Эти алгоритмы являются адаптивными методами семплирования по важности. Адаптивность заключается в том, что формулируется оптимизационная задача по минимизации дисперсии оценки, полученной методами семплирования по важности. Эта задача решается с помощью методов Франка-Вульфа (обычный и стохастические варианты).

Эти алгоритмы сходятся, однако стратегия выбора шага отличается у обычного и стохастического методов. В стохастических методах длина шага зависит явно от константы Липшица функции, оценка которой завышена в данной задаче. Поэтому обычный метод Франка-Вульфа сходится быстрее, чем стохастические варианты.

Дальнейшее исследование может быть связано с методов, использующих локальную константу Липшица, которая может оказаться сильно меньше, чем глобальная. Другая ветвь исследования связана с улучшением модели электрической сети: использованием «честной» модели переменного тока вместо линеаризованной модели постоянного тока.

6 Список литературы

- [1] Overbye Thomas J, Cheng Xu, Sun Yan. A comparison of the AC and DC power flow models for LMP calculations // System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on / IEEE. 2004. С. 9–pp.
- [2] Owen A. B., Maximov Y., Chertkov M. Importance sampling the union of rare events with an application to power systems analysis // ArXiv e-prints. 2017. .
- [3] Zimmerman R. D., Murillo-Sanchez C. E., Thomas R. J. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education // IEEE Transactions on Power Systems. 2011. Feb. Т. 26, № 1. С. 12–19.
- [4] Robert Christian P. Monte carlo methods. Wiley Online Library, 2004.
- [5] He Hera Y, Owen Art B. Optimal mixture weights in multiple importance sampling // arXiv preprint arXiv:1411.3954. 2014.
- [6] Adaptive Importance Sampling in General Mixture Classes / Olivier Cappé, Randal Douc, Arnaud Guillin [и др.]. 2008. 01. Т. 18. С. 447–459.
- [7] Robust Stochastic Approximation Approach to Stochastic Programming / Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan [и др.]. 2009. 01. Т. 19. С. 1574–1609.
- [8] Jaggi Martin. Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization // Proceedings of the 30th International Conference on Machine Learning / под ред. Sanjoy Dasgupta, David McAllester. Proceedings of Machine Learning Research № 1. Atlanta, Georgia, USA: PMLR, 2013. 17–19 Jun. С. 427–435. URL: <http://proceedings.mlr.press/v28/jaggi13.html>.
- [9] Hazan Elad, Luo Haipeng. Variance-Reduced and Projection-Free Stochastic Optimization // Proceedings of The 33rd International Conference on Machine Learning / под ред. Maria Florina Balcan, Kilian Q. Weinberger. Т. 48 из *Proceedings of Machine Learning Research*. New York, New York, USA: PMLR, 2016. 20–22 Jun. С. 1263–1271. URL: <http://proceedings.mlr.press/v48/hazana16.html>.

- [10] J. Reddi Sashank, Sra Suvrit, Póczos Barnabas [и др.]. Stochastic Frank-Wolfe methods for nonconvex optimization. 2016. 09.