

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ  
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)»  
ФИЗТЕХ-ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ  
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ  
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»

Кубентаева Самал

**Методы сэмплирования по важности  
для задачи оценки надёжности  
энергетических систем**

03.03.01 — Прикладные математика и физика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
(БАКАЛАВРСКАЯ ДИССЕРТАЦИЯ)

**Научный руководитель:**

к.ф-м.н.

Максимов Юрий Владимирович

Москва

2018

# Содержание

<b>1</b>	<b>Аннотация</b>	<b>2</b>
<b>2</b>	<b>Введение</b>	<b>3</b>
<b>3</b>	<b>Постановка задачи</b>	<b>5</b>
3.1	Нормальное распределение . . . . .	5
3.2	Стандартное нормальное распределение . . . . .	5
3.3	Энергетическая система . . . . .	7
<b>4</b>	<b>Методы</b>	<b>10</b>
4.1	Important Sampling . . . . .	10
4.2	Mixture IS . . . . .	11
4.3	ALOE . . . . .	12
4.4	mvtnorm . . . . .	13
<b>5</b>	<b>Новый подход</b>	<b>15</b>
5.1	Стохастический градиентный спуск . . . . .	15
5.2	Метод Франка-Вульфа . . . . .	17
<b>6</b>	<b>Эксперименты</b>	<b>18</b>
6.1	Результаты . . . . .	18
<b>7</b>	<b>Заключение</b>	<b>23</b>
	<b>Список литературы</b>	<b>24</b>

# 1 Аннотация

В настоящее время растёт доля электроэнергии, получаемой из возобновляемых источников, таких как ветер и солнце. Несмотря на экономическую эффективность возобновляемой энергетики, ветряная и солнечная генерации крайне нестабильны, что может вызвать перегрузку сети. Целью настоящей работы является оценка вероятности перегрузки сети в режимах линейной аппроксимации модели переменного тока. Предложены методы сэмплирования по важности для решения данной задачи, проведен их теоретический и экспериментальный анализ.

## 2 Введение

**Актуальность работы.** Люди начинают задумываться об окружающей среде всё больше и больше. Поэтому человечество переходит к использованию альтернативных источников энергии таких как, солнечные, ветряные электростанции, ГЭС, геотермальные станции и другие. На данный момент бурно развивается отрасль ветроэнергетики. Суммарная мощность ветрогенераторов превзошла суммарную мощность атомной энергии ещё в 2016 году. Энергия ветра является экологически чистым и экономически выгодным источником энергии.

В США значительное количество электроэнергии вырабатывается на ветроэнергетических установках (ВЭУ). В отличие от других видов источников, энергия ветра может вырабатывать неравномерную мощность, что приводит к сбоям системы, а именно, к недопустимо большим скачкам мощности. Они могут привести к нарушению работы системы и даже отключению электричества в штате. Надёжная система допускает малые колебания мощности. Наша задача состояла в том, чтобы уменьшить вероятность сбоев системы.

Мы будем оценивать эту вероятность для конкретного штата США, в котором возникла потребность в решении данной проблемы. Представим штат в виде графа  $G = (V, E)$ , где  $V$  — множество вершин,  $E$  — множество рёбер. Энергетическая система функционирует правильно, без сбоев, если вектор фаз (напряжения) не выходит из допустимой области  $D$ , которая представима в виде

$$D = \bigcup_{\{a,b\} \in E} (|\theta_a - \theta_b| \leq \tau_{ab})$$

где:

$|\theta_a - \theta_b|$  - разность фаз между вершинами  $a$  и  $b$ , которые соединены ребром  $\{a, b\} \in E$ ;  $\tau_{ab}$  - допустимая для данного ребра разность фаз, при которой ещё не произойдёт сбой в энергетической сети.

В реальном мире по электрической сети течёт переменный ток (АС), но мы будем описывать упрощенную модель сети, в которой ток будет постоянным (ДС). Такая модель описывается более простыми уравнениями, чем модель переменного

тока. Также линеаризованная модель постоянного тока имеет ряд преимуществ, таких как точное описание масштабных электрических сетей и высокая точность при высоковольтных режимах. Эта модель описана в работе [2]. Далее будем предполагать, что в короткие промежутки времени производство и потребление электроэнергии можно описать Нормальным распределением. Это распределение является довольно простым и наиболее известным на данный момент.

Сделав следующие предположения — линеаризация и Гауссовость, мы можем записать вероятность сбоя сети, как

$$\mu = P(\cup_{j=1}^J H_j)$$

где  $H_j$  — есть линейные ограничения на сеть, а  $J$  соответственно размер сети,  $j = 1, 2, \dots, J$ .

$$H_j = \{\mathbf{x}^\top \omega_j \geq \tau_j\}$$

по второму предположению  $\mathbf{x} \sim \mathcal{N}(\eta, \Sigma)$ .

Существует распределение Вейбулла, которое лучше описывает производимую и потребляемую мощности для ветровых источников. Возможно, в дальнейших работах над этой задачей мы сможем применить выборки из этого распределения. На данный момент известно слишком мало информации про распределение Вейбулла, поэтому в этой работе оно не может быть применено.

Работа разделена на четыре главные части. В первой, будет постановка задачи. Во второй части представлены существующие на данный момент методы решения данной задачи, которые были применены для улучшения результата. В третьей части, показан новый способ решения задачи, который использует методы SGD и Франка-Вульфа. В четвёртой части представлены полученные экспериментальные результаты на реальных данных из Matpowera и сравнение с методом ALOE.

## 3 Постановка задачи

### 3.1 Нормальное распределение

Предположим, что  $y \in \mathbb{R}^d$  — случайный вектор из  $\mathcal{N}(\eta, \Sigma)$ , ограничения имеют вид  $\hat{H}_j = \{y \mid u_j^\top y \geq \gamma_j\}$ . С помощью линейных преобразований получим случайный вектор  $x \in \mathbb{R}^d$ ,  $x \sim \mathcal{N}(0, I)$ , задаваемый уравнением

$$x = \Sigma^{-1/2}(y - \eta)$$

где  $\Sigma^{-1/2}$  существует, т.к.  $\Sigma$  положительно полуопределенная ковариационная матрица.

Теперь можем выразить ограничения  $\hat{H}_j$  через переменные  $x$ .

Подставим  $y = \eta - \Sigma^{-1/2}x$ , получим ограничения вида  $H_j = \{x \mid w_j^\top x \geq \tau_j\}$ ,  $\forall j \in \{1, 2, \dots, J\}$ , где

$$w_j = \frac{\Sigma^{-1/2}u_j}{\sqrt{u_j^\top \Sigma u_j}}$$
$$\tau_j = \frac{\gamma_j - u_j^\top \eta}{\sqrt{u_j^\top \Sigma u_j}}$$

Поэтому далее, без ограничения общности, будем считать  $x \sim \mathcal{N}(0, I)$  с ограничениями  $H_j \forall j \in \{1, \dots, J\}$ .

### 3.2 Стандартное нормальное распределение

Пусть  $x \in \mathbb{R}^d$  случайный вектор из  $\mathcal{N}(0, I)$ . Будем искать вероятность того, что  $x$  лежит вне многоугольника  $\mathcal{P}$ , который является допустимой зоной, где не происходит сбоев сети. Будем предполагать, что  $x \notin \mathcal{P}$  и для  $j = \{1, 2, \dots, J\}$  определим полупространства

$$H_j = \{x \mid w_j^\top x \leq \tau_j\}$$

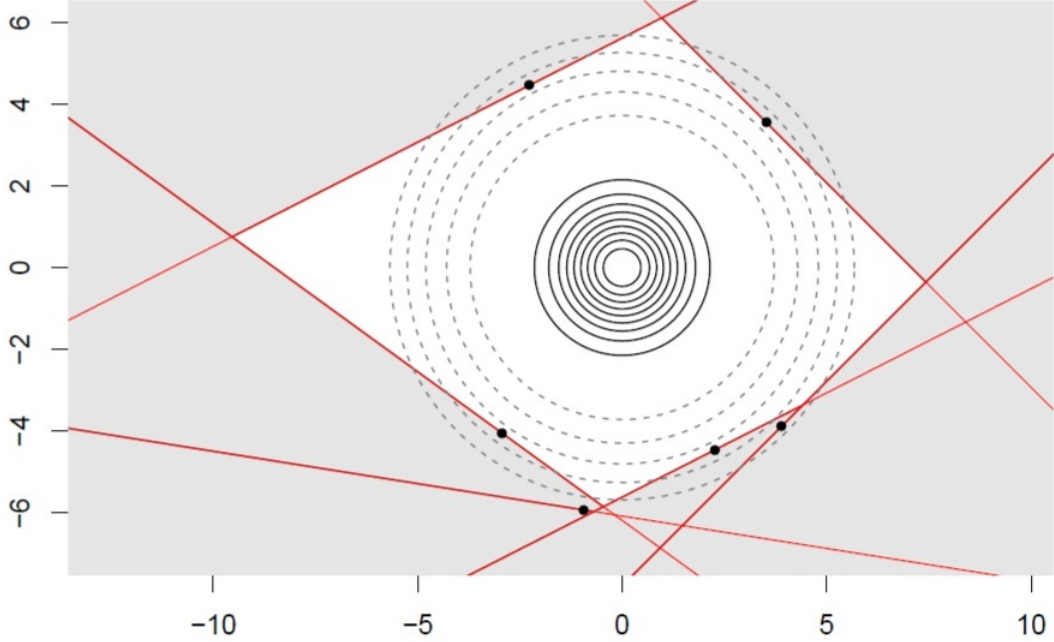


Рис. 1: многоугольник  $\mathcal{P}$ , заданный ограничениями  $H_j, j \in \{1, 2, \dots, J\}$ .

где  $\forall j: \tau_j \in \mathbb{R}$  и  $w_j \in \mathbb{R}^d, w_j^\top w_j = 1$ . Тогда  $\mathcal{P} = \bigcap_{j=1}^J H_j^c$ .

Вероятность, что  $x$  лежит вне многоугольника имеет вид:

$$\mu = P\{x \mid x \in H\}$$

где  $H = \bigcup_{j=1}^J H_j = \mathcal{P}^c$ . Множество  $\mathcal{P}$  выпуклое и не обязательно ограниченное,  $\tau_j > 0$ , т.к. мы ищем редкие события.

На рисунке 1 показан многоугольник  $\mathcal{P}$ , заданный ограничениями  $H_j, j \in \{1, 2, \dots, J\}$

Здесь, в четких окружностях находятся от 10% до 90% всех точек из  $\mathcal{N}(0, I)$ .

Пунктирные круги, содержат все точки, кроме тех, чья вероятность  $10^{-k}$ , где

$k \in \{3, 4, \dots, 7\}$ . Точки удовлетворяют условиям  $H_j$ . Нас интересуют редкие со-

бытия, т.е. такие  $\mathbf{x} \in \mathcal{N}(0, I)$ , что лежат в серой области.

### 3.3 Энергетическая система

Будем рассматривать энергетическую систему в виде графа  $G = (V, E)$ , где  $V$  — множество вершин графа, а  $E$  — множество рёбер.

Узлы графа могут быть трех типов. Первый,  $V_l$  (нагрузка) — это узлы, в которых нет генераторов, для них известны величины реальной и реактивной мощности. Второй,  $V_g$  (генераторы) — это узлы с одним генератором, в которых известны реальная мощность и величина напряжения  $|V|$ . Третий тип,  $V_s$  балансирующие узлы, в которых известны величина напряжения и фазы. Опишем энергетическую систему опираясь на источники [1] и [8].

Рассмотрим уравнения потока мощности:

$$P_k = \sum_{j=1}^N |V_k||V_j|(G_{kj} \cos(\theta_k - \theta_j) + B_{kj} \sin(\theta_k - \theta_j))$$

$$Q_k = \sum_{j=1}^N |V_k||V_j|(G_{kj} \sin(\theta_k - \theta_j) - B_{kj} \cos(\theta_k - \theta_j))$$

Введем три предположения относительно высоковольтных энергетических систем:

- сопротивление линий передач значительно ниже реактивного сопротивления;
- разность фаз на соседних узлах мала (обычно, разность фаз на соседних узлах не превышает 10-15 градусов, крайне редко составляет 30 градусов);
- величины  $|V_k|$  и  $|V_j|$  близки к 1.0

Тогда получим следующий вид уравнений баланса мощности:

$$P_k = \sum_{j=1}^N B_{kj}(\theta_k - \theta_j)$$

$$Q_k = -b_k + \sum_{j=1, j \neq k}^N b_{kj}(|V_k| - |V_j|)$$



Можно заметить, что реальный поток мощности в сети значительно больше потока реактивной мощности.

$$P_{kj} \gg Q_{kj}$$

Таким образом, при оценке перегрузки сети, разумно смотреть только на реальные составляющие мощности. Поэтому в нашей задаче оптимизации, мы будем рассматривать только уравнения реального потока мощности.

$$\mathbf{p} = B\theta$$

Мы рассматриваем линейризованную модель постоянного тока (DC), поэтому уравнение баланса мощности имеет вид:

$$\mathbf{p} = B\theta$$

где  $B$  матрица Лапласа,  $\mathbf{p}$  вектор реальных мощностей,  $\theta$  разность фаз. Следовательно, перепишем:

$$\theta = B^+ \mathbf{p}$$

где  $B^+$  — псевдообратная матрица.

Так как мы рассматриваем DC модель, то имеем следующие ограничения на мощности и фазы:

$$|\theta_k - \theta_j| \leq \theta_{kj} \quad (1)$$

$$\underline{p^S} \leq p^S \leq \overline{p^S} \quad (2)$$

$$\underline{p^g} \leq p^g \leq \overline{p^g} \quad (3)$$

Выразим все ограничения через  $p^g$ . Для этого представим  $B = [B_l, B_g, B_s]$  и  $D_{kj} \in \{-1, 0, 1\}$ , где  $D_{kj} = -1$ , если  $k$ -ое ребро заканчивается на  $j$ -м узле,  $D_{kj} = 1$ , если  $k$ -ое ребро начинается на  $j$ -м узле, во всех остальных случаях  $D_{kj} = 0$ . Тогда

получим ограничения на фазу следующего вида:

$$(DB_g^+ - DB_s^+ 1_g^\top) p^g \leq \bar{\theta} - (DB_l^+ - DB_s^+ 1_g^\top) p^l \quad (4)$$

$$(DB_s^+ 1_g^\top - DB_g^+) p^g \leq \bar{\theta} + (DB_l^+ - DB_s^+ 1_g^\top) p^l \quad (5)$$

$$-\overline{p^S} - 1_l^\top \leq 1_g^\top p^g \leq -\underline{p^S} - 1_l^\top \quad (6)$$

Так получили все нужные ограничения для решения задачи, при условии что  $p^g \sim \mathcal{N}(\eta^g, \Sigma^g)$ .

## 4 Методы

### 4.1 Important Sampling

Применим идеи метода выборки по важности (IS) для оценки вероятности редких событий [10]. IS является лучшим способом поиска выборки из области редких событий.

Рассмотрим следующую задачу оценки интеграла:

$$\mu = \mathbb{E}f(X) = \int f(x)p(x)dx \quad (7)$$

где  $X \sim p$  случайная величина из  $\mathbb{R}^d$ ,  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ .

Стандартный метод Монте-Карло оценивает интеграл, беря независимые одинаково распределенные (IID) выборки  $X_1, X_2, \dots, X_n \sim p$  и используя

$$\hat{\mu}_n^{MC} = \frac{1}{n} \sum_{i=1}^n f(X_i)$$

Для этой оценки  $\mathbb{E}\hat{\mu}_n^{MC} = \mu$  и

$$\text{Var}(\hat{\mu}_n^{MC}) = \frac{1}{n} \text{Var}_{X \sim p}(f(X)) = \frac{1}{n} \left( \int f^2(x)p(x)dx - \mu^2 \right)$$

Воспользуемся идеей IS, берем IID выборку  $X_1, X_2, \dots, X_n \sim q$ , где  $q$  — распределение выборки, удовлетворяет уловию  $q(x) > 0$  каждый раз, когда  $f(x)p(x) \neq 0$ . Оценка будет следующего вида

$$\hat{\mu}_n^{IS} = \frac{1}{n} \sum_{i=1}^n f(X_i) \frac{p(X_i)}{q(X_i)} \quad (8)$$

Для которой также  $\mathbb{E}\hat{\mu}_n^{IS} = \mu$  и

$$\text{Var}(\hat{\mu}_n^{IS}) = \frac{1}{n} \text{Var}_{X \sim q} \left( f(X) \frac{p(X)}{q(X)} \right) = \frac{1}{n} \left( \int f^2(x) \frac{p^2(X)}{q(X)} dx - \mu^2 \right) \quad (9)$$

Далее будем решать задачу минимизации полученной дисперсии, принимая во внимание факт того, что мы ищем оценку редких событий. То есть важно получить выборки вне многогранника. Для этого применим идеи из статьи [6].

## 4.2 Mixture IS

Главным плюсом этого метода является умение работать с редкими событиями. Mixture IS определяет распределение выборки, из которой можно строить выборку из области редких событий. Берем  $q_j$  из смеси распределений  $\{q_1, \dots, q_J\}$  с вероятностью  $\alpha_j$ ,  $\alpha_j > 0$  и  $\sum_{j=1}^J \alpha_j = 1$ . Тогда распределение выборки будет следующего вида

$$q_\alpha(x) = \sum_{j=1}^J \alpha_j q_j(x) \quad (10)$$

Подставим полученное распределение (10) в (8), получим следующую оценку

$$\mu_\alpha = \frac{1}{n} \sum_{i=0}^n \frac{f(X_i)p(X_i)}{q_\alpha(X_i)} = \frac{1}{n} \sum_{i=0}^n \frac{f(X_i)p(X_i)}{\sum_{j=1}^J \alpha_j q_j(X_i)} \quad (11)$$

Аналогично для дисперсии (9)

$$\mathbf{Var}(\mu_\alpha) = \frac{1}{n} \left( \int \frac{f^2(x)p^2(x)}{\sum_{j=1}^J \alpha_j q_j(x)} dx - \mu^2 \right) \quad (12)$$

### Обозначения:

- $H_j = \{\mathbf{x} \mid w_j^\top \mathbf{x} \leq \tau_j\}$  — ограничения;
- $f(x) = H_j(x) = \mathbb{I}(x \in H_j)$ ;
- $P_j = P(\mathbf{x} \in H_j)$  — вероятность того, что  $x$  удовлетворяет ограничению  $H_j$ ;
- $p(x)$  — функция плотности вероятности  $\mathcal{N}(0, I)$ ;
- $S(x_i) = \sum_{j=1}^J H_j$  — количество произошедших редких событий;
- $T_s = P(S = s), s \in \{0, 1, \dots, J\}$

- $\bar{\mu} = \sum_{j=1}^J P_j$

Нужно определить набор распределений  $\{q_j\}_{j=1}^J$  и значения весов  $\alpha_j$ , чтобы использовать метод Mixture IS. Для этого, воспользуемся выводами полученными в статье [1].

### 4.3 ALOE

Пусть для  $j \in \{1, 2, \dots, J\}$   $q_j = \{\mathcal{L}(\mathbf{x} \mid w_j^\top \mathbf{x} \leq \tau_j)\}$  и берем  $q_0 = p = \mathcal{N}(0, I)$ , с  $P_0 = 1$  и  $H_0(x) = 1$ . Функция плотности вероятности для  $j > 0$  будет

$$q_j(x) = \frac{p(\mathbf{x})H_j(\mathbf{x})}{P_j} \quad (13)$$

Выбирая точки таким образом будем получать только те, что находятся вне многоугольника, как нам и нужно. Можем записать новую формулу для оценки и её дисперсии, где  $\mathbf{x}_i \sim q_\alpha = \sum_{j=0}^J \alpha_j q_j$ :

$$\mu_\alpha = \frac{1}{n} \sum_{i=1}^n \frac{H_{1:J}(\mathbf{x}_i)p(\mathbf{x}_i)}{\sum_{j=0}^J \alpha_j q_j(\mathbf{x}_i)} = \frac{1}{n} \sum_{i=1}^n \frac{H_{1:J}(\mathbf{x}_i)}{\sum_{j=0}^J \alpha_j H_j(\mathbf{x}_i)P_j^{-1}},$$

$$\mathbf{Var}(\mu_\alpha) = \frac{1}{n} \left( \int \frac{H_{1:J}(\mathbf{x}_i)p^2(\mathbf{x}_i)}{\sum_{j=0}^J \alpha_j q_j(\mathbf{x}_i)} dx - \mu^2 \right) = \frac{1}{n} \left( \int \frac{H_{1:J}(\mathbf{x}_i)p(\mathbf{x}_i)}{\sum_{j=0}^J \alpha_j H_j(\mathbf{x}_i)P_j^{-1}} dx - \mu^2 \right). \quad (14)$$

Вектор весов смеси  $\alpha$  в Mixture IS пропорционален вероятностям  $P_j, \forall j \in \{1, 2, \dots, J\}$ , т.е.

$$\alpha_{ALOE} = \frac{P_j}{\sum_{j=1}^J P_j}. \quad (15)$$

Теперь при  $\alpha^* = \alpha_{ALOE}$ :

$$\mu_{\alpha^*} = \frac{\bar{\mu}}{n} \sum_{i=1}^n \frac{H_{1:J}(\mathbf{x}_i)}{\sum_{j=1}^J H_j(\mathbf{x}_i)} = \frac{\bar{\mu}}{n} \sum_{i=1}^n \frac{1}{S(\mathbf{x}_i)}, \quad \mathbf{x}_i \sim q_{\alpha^*}, \quad (16)$$

т.к. для  $\mathbf{x}_i \sim q_{\alpha^*}$  всегда выполняется  $H_{1:J} = 1$ . Заметим, что  $S(\mathbf{x}_i)^{-1}$  может принимать значения от 1 до  $J$ , поэтому можем записать следующую оценку на  $\mu$ :

$$\frac{\bar{\mu}}{J} \leq \mu \leq \bar{\mu}.$$

**Теорема 1.** Пусть  $\mu_{\alpha} = \mu_{\alpha^*}$ . Тогда  $\mathbb{E}\mu_{\alpha} = \mu$  и

$$\mathbf{Var}(\mu_{\alpha}) = \frac{1}{n} \left( \bar{\mu} \sum_{s=1}^J \frac{P(S=s)}{s} - \mu^2 \right) \leq \frac{\mu(\bar{\mu} - \mu)}{n}$$

**Теорема 2.** Пусть  $\mu_{\alpha} = \mu_{\text{ALOE}}$ , где  $T_s = P(S=s)$ . Тогда

$$\mathbf{Var}(\mu_{\alpha}) = \frac{\mu^2}{n} \left( \left( \sum_{s=1}^J s \frac{T_s}{\mu} \right) \left( \sum_{s=1}^J s^{-1} \frac{T_s}{\mu} \right) - 1 \right) \leq \frac{\mu^2(J + J^{-1} - 2)}{4n}.$$

Верхняя граница достигается при  $S \sim \mathcal{U}(1, J)$

Таким образом, метод ALOE не учитывает геометрию многогранника, что приводит к проблемам при организации выборки из распределения  $q_{\alpha^*}$ . Так же существуют случаи, при которых оценка дисперсии достигает своей верхней границы. Как показала практика, именно такие случаи чаще всего встречаются в реальной жизни.

## 4.4 mvtnorm

Был рассмотрен пример, сравнивающий ALOE с mvtnorm из пакета R mvtnorm [4]. Данный пакет работает на задачах с большой размерностью и может использовать некоторые свойства Гауссовского распределения. Опишем, mvtnorm опираясь на статью [5]. Программа вычисляет

$$P(\mathbf{a} \leq \mathbf{y} \leq \mathbf{b}) = P(a_i \leq y_i \leq b_i), \quad i = 1, 2, \dots, d$$

где  $\mathbf{y} \sim \mathcal{N}(\eta, \Sigma)$ ,  $-\infty \leq a_i \leq b_i \leq \infty$ ,  $i = 1, 2, \dots, d$ .

Используем её для нахождения  $\mu = P(\sum_{j=1}^J \mathbb{I}\{w_j^\top \mathbf{x} \geq \tau_j\} > 0)$  для  $\mathbf{x} \sim \mathcal{N}(0, I)$ .

$$1 - \mu = P(\Omega^\top \mathbf{x} \leq) = P(\mathbf{y} \leq), \quad \mathbf{y} \sim \mathcal{N}(0, \Omega^\top \Omega)$$

Программа может работать с размерностью более чем 1000, что в нашем случае соответствует  $J = 1000$  линейных ограничений.

## 5 НОВЫЙ ПОДХОД

Будем адаптивно подбирать веса  $\alpha_j$ ,  $j = 1, 2, \dots, J$  для Mixture IS. Решаем следующую задачу оптимизации:

$$\mathbf{Var}(\mu_\alpha) = \frac{1}{n} \left( \int \frac{H_{1:J}(\mathbf{x}_i)p(x_i)}{\sum_{j=1}^J \alpha_j H_j(\mathbf{x}_i)P_j^{-1}} dx - \mu^2 \right) \rightarrow \min_{\alpha \in S_0} \quad (17)$$

где  $S_0 = \{(\alpha_1, \alpha_2, \dots, \alpha_J) \mid \forall j \alpha_j > 0, \sum_{j=1}^J \alpha_j = 1\}$  — симплекс.

**Теорема 3.** Дисперсия  $\mu_\alpha$ , заданная формулой (14), является выпуклой функцией по  $\alpha$  на симплексе  $S_0$ .

Эта теорема была доказана в [6].

Введем предположение, что  $\forall j \alpha_j \geq \varepsilon > 0$  для того, чтобы гарантировать стабильность и сходимость метода. Этот подход поможет выбирать разумное количество точек из распределения  $q_\alpha$ , при этом учитывая вклад всех ограничений.

### 5.1 Стохастический градиентный спуск

Этот метод является простым и при этом очень эффективным в больших задачах. Обычно, решает задачи следующего вида:

$$\min_{x \in \Omega} F(x) = \begin{cases} \mathbb{E}_z f(x, z) & \text{(стохастическая)} \\ \frac{1}{n} \sum_{i=1}^n f_i(x) & \text{(конечная сумма)} \end{cases} \quad (18)$$

Приведем  $F(\alpha)$  к стохастическому виду:

$$F(\alpha) = n \mathbf{Var} \mu_\alpha + \mu^2 = \int \frac{H_{1:J}(x)p(x)}{\sum_{j=0}^J \alpha_j H_j(x)P_j^{-1}} dx = \mathbb{E}_{q_\alpha} \frac{H_{1:J}(x)}{\left( \sum_{j=0}^J \alpha_j P_j^{-1} H_j(x) \right)^2} \quad (19)$$



Будем решать следующую задачу минимизации:

$$\mathbb{E}_{q_\alpha} \frac{H_{1:J}(x)}{\left(\sum_{j=0}^J \alpha_j P_j^{-1} H_j(x)\right)^2} \longrightarrow \min_{\alpha \in S}. \quad (20)$$

где  $S = \{(\alpha_1, \alpha_2, \dots, \alpha_J) \mid \forall j \alpha_j \geq \varepsilon, \sum_{j=1}^J \alpha_j = 1\}$

Стохастический градиент будет иметь вид:

$$\nabla F(x) = \mathbb{E}_{q_\alpha} \frac{-\mathbb{I}(x \in H) P_j^{-1} H_j(x)}{\left(\sum_{j=0}^J \alpha_j P_j^{-1} H_j(x)\right)^3}$$

Будем применять алгоритм из [10] с единственным изменением, оптимизация будет проходить по параметру  $\alpha$ , а не по  $x$ .

---

### Algorithm 1 SGD ( $\alpha^{(0)}, T, C$ )

---

**Require:**  $\alpha^{(0)}$ , number of iterations  $T$ , step size  $C$ ;

**Ensure:** Sequence  $\{\alpha^{(t)}\}_{t=0}^{T-1}$ , estimation  $\mu_\alpha$ ;

**for**  $t = 0, \dots, T - 1$  **do**

$x_t \sim q_{\alpha^{(t)}}$ ;

update estimation  $\mu_\alpha = \frac{1}{t} \sum_{i=0}^t \frac{H_{1:J}(x_i)}{\sum_{j=1}^J \alpha_j^{(i)} P_j^{-1} H_j(x_i)}$ ;

Compute  $\nabla_t F(x) = \mathbb{E}_{q_{\alpha^{(t)}}} \frac{-\mathbb{I}(x \in H) P_t^{-1} H_t(x)}{\left(\sum_{j=0}^J \alpha_j P_j^{-1} H_j(x)\right)^3}$ ;

Step size: in case of fixed size policy  $C_t = C$ , otherwise  $C_t = C/\sqrt{t}$ ;

Step  $\alpha^{(t+1)} = \alpha^{(t)} - C_t \nabla_t F(x)$ ;

$\alpha^{(t+1)} = \prod_{S(\alpha^{(t+1)})}$ .

---

Стохастический градиентный метод сходится к оптимальному значению и его качество сильно зависит от констант  $C$  и  $\varepsilon$ . При слишком большом шаге метод может не сойтись. Если  $\alpha = \varepsilon$  и значение  $\varepsilon$  мало, то скорость сходимости метода будет низкой. Однако, при больших значениях  $\varepsilon$  мы можем потерять оптимальные точки, что плохо скажется на результате.

## 5.2 Метод Франка-Вульфа

Приведем задачу к виду конечных сумм, чтобы решить её методом Франка-Вульфа [7]. Имея выборку  $x_1, \dots, x_n \sim q_\alpha$ , можем записать задачу (20) в виде:

$$F(\alpha) = \frac{1}{n} \sum_{i=1}^n \frac{H_{1:J}(x_i)}{\left(\sum_{j=0}^J \alpha_j H_j(x_i) P_j^{-1}\right)^2} \longrightarrow \min_{\alpha \in S} \quad (21)$$

Будем применять стандартный алгоритм Франка-Вульфа:

---

**Algorithm 2** Frank-Wolfe ( $\alpha^{(0)}, T$ )

---

**Require:**  $\alpha^{(0)}$ , number of iterations  $T$ ;

**Ensure:** Sequence  $\{\alpha^{(t)}\}_{t=0}^T$ ;

  Compute  $g_0 = \frac{1}{n} \sum_{i=1}^n \nabla f_i(\alpha_0^i)$ ;

**for**  $t = 1, \dots, T$  **do**

    Compute  $v_t = \arg \max_{v \in S} \langle v, \nabla F(\alpha^{(t-1)}) \rangle$ ;

    Compute update direction  $d_t = v_t \alpha^{(t-1)}$ ;

$\gamma_t = \frac{2}{t+1}$

    step  $\alpha^{(t+1)} = \alpha^{(t)} + \gamma_t d_t$

---

Метод Франка-Вульфа не решает начальную задачу, зато может хорошо приблизить функцию к  $\hat{F}^*$  за не большое число итераций. Выполнено  $\mathbb{E}_{q_\alpha}(\hat{F}^*) = F^*$  для  $\hat{F}^*$ . Метод вычисляет полный градиент функции за  $O(NJ)$ , то есть при увеличении  $N$  будет улучшена точность дисперсии, но также повысится сложность операции.

## 6 Эксперименты

В качестве данных были использованы пакеты из MATPOWER, которые близки к реальным значениям мощности в энергосистеме. Используются функции и описание данных взятых из [9]. Эксперименты проводились на пакетах case1354pegase, case2869pegase и case9241pegase, описание которых также взято из [3]. Аналогично [1], смодлировано нарушение фазовых ограничений сети.

Case	$N$	$N_R$	$J$
case1354pegase	1354	260	3464
case2869pegase	2869	510	8146
case9241pegase	9241	1445	29210

Таблица 1: Matpower cases

где  $N$  — количество узлов,  $N_R$  — узлы с генераторами,  $J = 2N - 2N_R + 2$  — количество ограничений.

### 6.1 Результаты

Применены методы SGD и Франк-Вульф, описанные в предыдущем разделе. Было отмечено, что в большинстве случаев, существует одно редкое событие, которое имеет большую вероятность относительно других редких событий. Например, такой случай возник в пакете case1354pegase. Такой же результат был получен в [1].

SGD был запущен со следующими параметрами:  $\alpha^{(0)} = \alpha_{ALOE}$ , шаг метода равен  $\sqrt{2}/G$ , где  $G \sim (\bar{\mu}/\underline{\mu})^2$ . Напомним, что  $\bar{\mu} = \sum_{j=0}^J P_j$ ,  $\underline{\mu} = \max_{j \in \{0,1,J\}} P_j$ .

Метод Франка-Вульфа использует параметр  $\alpha^{(0)} = \alpha_{ALOE}$ , так как это лучшее значение параметра  $\alpha$  известное на данный момент.

Для обоих методов  $\varepsilon = \frac{0.1}{J+1}$ , так как из метода Mixture IS будем считать что вклад каждого распределения смеси будет не меньше 10% от общей доли. Но оставшиеся 90% метод сам будет перераспределять, то есть такое предположение не мешает работе алгоритма.

Были получены следующие результаты на разных пакетах данных:

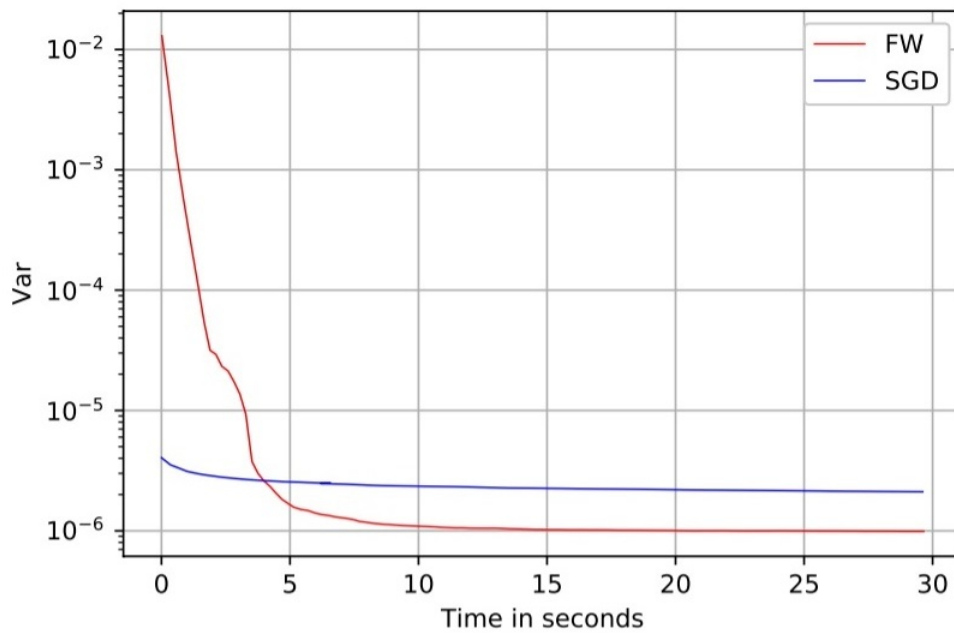


Рис. 2: case1354pegase

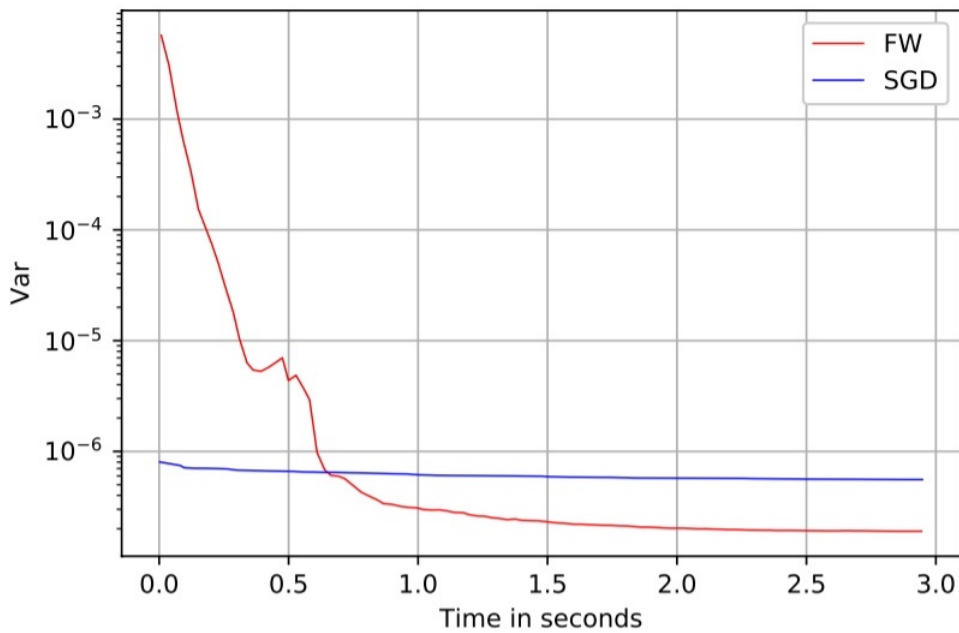


Рис. 3: case2869pegase

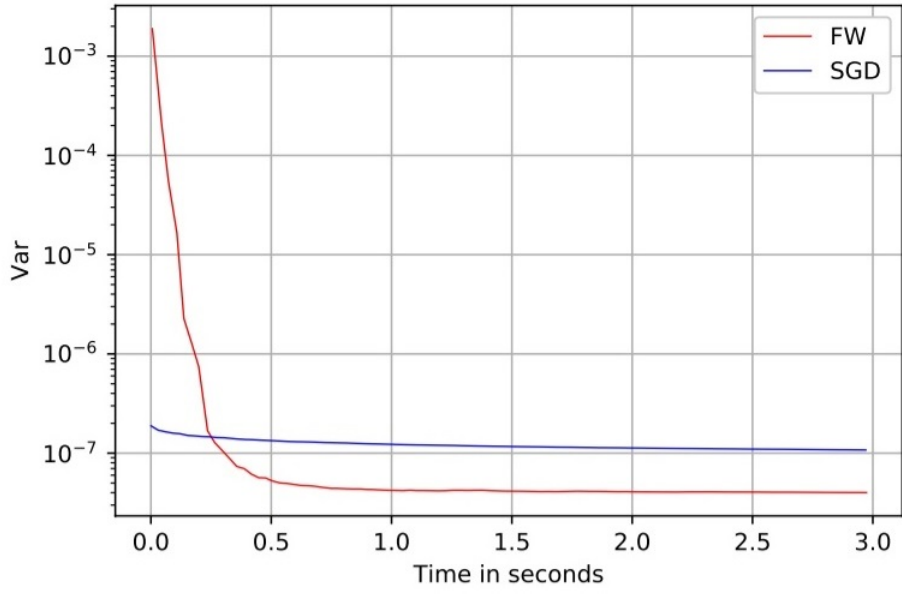


Рис. 4: case9241pegase

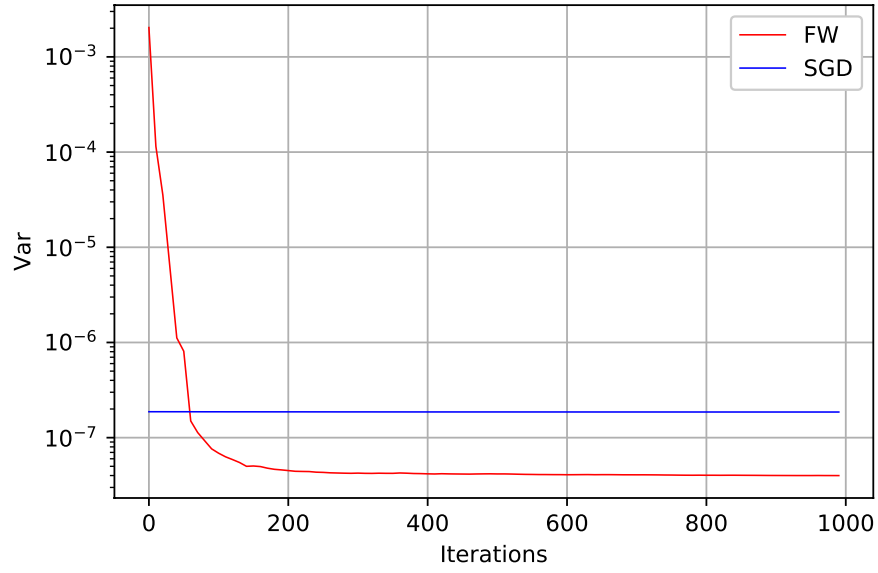


Рис. 5: case1354pegase

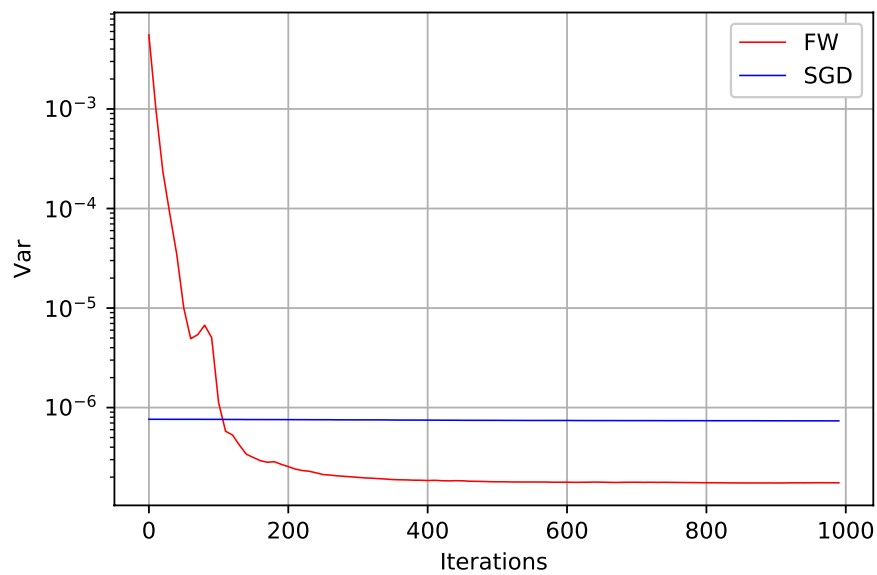


Рис. 6: case2869pegase

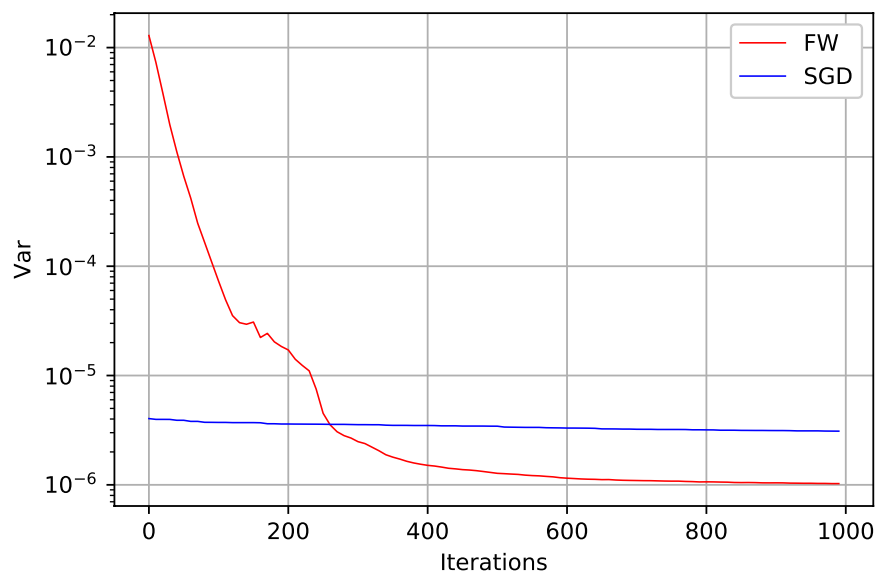


Рис. 7: case9241pegase

Как видно из полученных графиков оба метода сошлись к оптимуму. Хотя SGD не сделал больших успехов, по сравнению с методом Франка-Вульфа, который меньше чем за секунду сошелся к оптимальному значению.

Как видно, Франк-Вульф дает более точный результат чем SGD, что вполне естественно, так как он подбирает направление спуска и пересчитывает полный градиент функции. Также размерность задачи слабо влияет на скорость сходимости метода, так как на пакете с количеством ограничений большим на порядок, он дал такие же хорошие результаты.

Метод Франка-Вульфа сделал около 200 итераций на пакетах case1354pegase и case2869pegase за время меньше 1 секунды. На данных case9241pegase сделано около 400 итераций, что так же привело к точным результатам. Таким образом, метод Франка-Вульфа является эффективным при решении поставленной задачи. Приведем пример сравнения методов ALOE, SGD и Франка-Вульфа в следующей таблице, где для ограничений (4) и (5) значение  $\bar{\theta} = \pi/6$ :

100000 выборка			
	$\mu_\alpha$	$\mathbf{Var}\mu_\alpha$	$\frac{\sqrt{\mathbf{Var}\mu_\alpha}}{\mu_\alpha}$
ALOE	$2.31 \cdot 10^{-4}$	$3.5 \cdot 10^{-13}$	0.003
SGD	$2.30 \cdot 10^{-4}$	$7.2 \cdot 10^{-13}$	0.003
Frank-Wolfe	$2.31 \cdot 10^{-4}$	$1.1 \cdot 10^{-13}$	0.001
Lower bound	$2.30 \cdot 10^{-4}$		
Upper bound	$5.70 \cdot 10^{-4}$		

Таблица 2: Оценки на данных из case2869pegase

Как видно из данной таблицы SGD дает не лучший результат, но его значения имеют тот же порядок, что и метод ALOE. Данный результат работы SGD был предсказуем, так как при очень малом  $\varepsilon$ , значение  $G^2$  большое, что плохо влияет на сходимость метода. Метод Франка-Вульфа сошелся к тому же результату, что и метод ALOE и дал значительно лучший результат по сравнению с SGD.

Метод Франка-Вульфа превзошел метод ALOE, так как он учитывает геометрию многогранника. В случаях достижения верхней границы оценки ALOE, Франк-Вульф будет значительно эффективнее.

## 7 Заключение

Были рассмотрены методы Important sampling, Mixture IS и ALOE для улучшения оценки надежности энергетической системы. Предложен новый метод решения данной задачи, который заключался в оптимизации параметров сети, таких как веса компонент смеси.

Для этого применены методы SGD и Франка-Вульфа. SGD является методом решающим стохастические задачи оптимизации, а для переформулированной задачи в виде конечных сумм применили метод классического Франка-Вульфа.

SGD при константном выборе шага может не сойтись, а при шаге  $C_t = C/\sqrt{t}$  сходимость слишком медленная. Как было показано в экспериментах, скорость сходимости метода была низкой.

В отличие от SGD, метод Франка-Вульфа дал хорошие результаты и показал быструю сходимость к оптимуму (менее чем за секунду), даже на масштабных данных. Однако, метод будет работать медленнее, если потребуется большое количество точек, по которым производится оценка, так как для генерации такой выборки может уйти много времени. Также эффективность метода может снизиться при большом количестве ограничений  $J$ , т.к. на подсчет одной итерации понадобится больше времени.

В дальнейших работах над этой задачей, можно улучшить качество работы данных методов, путем улучшения входных параметров и констант. Также можно решать более общую задачу для AC модели.



## Список литературы

- [1] A. B. Owen Y. Maximov, Chertkov M. Importance sampling the union of rare events with an application to power systems analysis. — arXiv preprint arXiv:1710.06965, October 2017.
- [2] B. Stott J. Jardim, Alsac. O. DC power flow revisited. — 24(3):1290–1300 : IEEE Transactions on Power Systems, Aug 2009.
- [3] C. Jozs S. Fliscounakis J. Maeght, Panciatici P. MAC Power Flow Data in MATPOWER and QCQP format: iTesla, RTE Snapshots, and PEGASE. — arXiv:1603.01533, Mar 2016.
- [4] Genz A. Numerical computation of rectangular bivariate and trivariate normal and t probabilities. — 14(3):251–260 : Statistics and Computing, 2004.
- [5] Genz A., Bretz F. Computation of Multivariate Normal and t Probabilities. — Berlin : Springer-Verlag, 2009.
- [6] He H. Y., Owen. A. B. Optimal mixture weights in multiple importance sampling. — arXiv preprint arXiv:1411.3954, November 2014.
- [7] Jaggi. Martin. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. — volume 28 of Proceedings of Machine Learning Research, pages 427–435, Atlanta, Georgia : Proceedings of the 30th International Conference on Machine Learning, Jun 2013.
- [8] Kenneth Van den Bergh Erik Delarue, D’haeseleer William. DC power flow in unit commitment models. — Energy and Environment, May 2014.
- [9] R. D. Zimmerman C. E. Murillo-Sanchez, Thomas. R. J. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. — IEEE Transactions on Power Systems, 26(1):12–19, Feb 2011.
- [10] Ryu E. K., Boyd S. P. Adaptive Importance Sampling via Stochastic Convex Programming. — arXiv preprint arXiv:1412.4845, December 2014.