Этот документ служит математическим обоснованием метода диффузии для проекта по супер резолюции полей погоды.

# 1 Вступление

В статьях [1][4] предложен метод параметризации процесса дифузии, показавший хорошие результаты в задаче генерации изображений. Этот метод в последующем был расширен [2] на задачу условной генерации данных, что позволяет решать в том чилсле задачу супер резолюции. Предлагается дальше расширить этот метод, добавив в него возможность регулеризировать получающиеся изображения при помощи дополнительных перобразовний исходного изображения некоторыми функциями.

Предпологается, что данная модель будет использована для полей погоды. Для них хороший результат показывает регулеризвация нормы матрицы дискретного спектра фурье [3]. В связи с этим, предлагается рассмотреть частный случай, где в качестве функции выступает дискретное преобразование фурье.

# 2 Совместное распределение

Прямой процесс диффузии характерезуется условным распределнием

$$q(x_t|x_{t-1}) = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon$$

Где $\epsilon \sim \mathcal{N}(0, I)$. Рассмотрим расширенный вектор $x_t^*$, добавив к вектору $x_t$ его отображение преобразованием $f$.

$$x_t^* = (x_t, f(x_t))$$

Предположим, что $f$ - это линейная биекция и выполняется

$$f(\epsilon) \sim \mathcal{N}(0, I_{n'}), \epsilon \sim \mathcal{N}(0, I_n) \tag{1}$$

Например в качестве $f$ - можно взять линейное, ортогональное преобразованние

$$f(x) = Ax, AA^T = I, A \in \mathbb{R}^{n\times n}$$

Тогда случайную переменную $f(x_t)$ можно записать в виде

$$\begin{aligned} f(x_t) &= f(\sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon) \\ &= \sqrt{1-\beta_t}f(x_{t-1}) + \sqrt{\beta_t}f(\epsilon) \end{aligned}$$

где $\epsilon \sim \mathcal{N}(0, I)$. Пользуясь (1)

$$q(f(x_t)|x_{t-1}) = \sqrt{1-\beta_t}f(x_{t-1}) + \sqrt{\beta_t}\epsilon = \mathcal{N}(\sqrt{1-\beta_t}f(x_{t-1}), \beta_t I)$$

Что в точности повторяет прямой процесс диффузии $q(x_t|x_{t-1})$ для преобразованной переменной $f(x_t)$. Обозначим $f_t = f(x_t)$. Тогда дифузия вектора $(x_0, f_0)$ эквивалентна

$$q(x_T, f_T, .., x_0, f_0) = q(x_0, f_0) \prod_{t=1}^{T} q(x_t, f_t|x_{t-1}, f_{t-1})$$

с переходами

$$\begin{aligned} q(x_t, f_t|x_{t-1}, f_{t-1}) &= q(x_t|x_{t-1})q(f_t|f_{t-1}) \\ q(x_t|x_{t-1}) &= \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \sqrt{\beta_t}I) \\ q(f_t|f_{t-1}) &= \mathcal{N}(\sqrt{1-\beta_t}f_{t-1}, \sqrt{\beta_t}I) \\ q(x_0, f_0) &= q(x_0)q(f_0),\, q(f_0) = q(f(x_0)) \end{aligned}$$

Что бы параметризовать обратный процесс, скажем что

$$\begin{aligned} p(x_T, f_T, .., x_0, f_0) &= p(x_0, f_0) \prod_{t=1}^{T} p(x_{t-1}, f_{t-1}|x_t, f_t) \\ p_\theta(x_{t-1}, f_{t-1}|x_t, f_t) &= p_\theta(x_{t-1}|x_t)p_\theta(f_{t-1}|f_t) \\ p_\theta(x_T, f_T) &= p_\theta(x_T)p_\theta(f_T) \\ p_\theta(x_T) &= p_\theta(f_T) = \mathcal{N}(0, I) \end{aligned}$$

Тогда следуя выкладкам из [1] можно сказать, что вариационная нижняя оценка выражается в виде

$$
\begin{aligned}
L = \mathbb{E}_q \Bigg[ & -\log p(x_T, f_T) - \sum_{t>1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})} - \log \frac{p_\theta(x_0|x_1)}{q(x_1|x_0)} \\
& - \sum_{t>1} \log \frac{p_\theta(f_{t-1}|f_t)}{q(f_t|f_{t-1})} - \log \frac{p_\theta(f_0|f_1)}{q(f_1|f_0)} \Bigg] \\
= \mathbb{E}_q \Bigg[ & -\log p(x_T, f_T) - \sum_{t>1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} \frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} - \log \frac{p_\theta(x_0|x_1)}{q(x_1|x_0)} \\
& - \sum_{t>1} \log \frac{p_\theta(f_{t-1}|f_t)}{q(f_{t-1}|f_t, f_0)} \frac{q(f_{t-1}|f_0)}{q(f_t|f_0)} - \log \frac{p_\theta(f_0|f_1)}{q(f_1|f_0)} \Bigg] \\
= \mathbb{E}_q \Bigg[ & -\log \frac{p(x_T)}{q(x_T|x_0)} - \sum_{t>1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} - \log p_\theta(x_0|x_1) \\
& -\log \frac{p(f_T)}{q(f_T|f_0)} - \sum_{t>1} \log \frac{p_\theta(f_{t-1}|f_t)}{q(f_{t-1}|f_t, f_0)} - \log p_\theta(f_0|f_1) \Bigg] \\
= \mathbb{E}_q \Bigg[ & -\log \frac{p(x_T)p(f_T)}{q(x_T, f_T|x_0)} - \sum_{t>1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} \\
& - \sum_{t>1} \frac{p_\theta(f_{t-1}|f_t)}{q(f_{t-1}|f_t, f_0)} - \log p_\theta(x_0, f_0|x_1, f_1) \Bigg]
\end{aligned}
$$

Обозначим

$$
L_{t-1} = -\log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} - \log \frac{p_\theta(f_{t-1}|f_t)}{q(f_{t-1}|f_t, f_0)}
$$

Для распределения $q(x_{t-1}|x_t, x_0)$ имеем

$$
\begin{aligned}
q(x_{t-1}|x_t, x_0) = &\quad \mathcal{N}(\tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I) \\
\tilde{\mu}_t(x_t, x_0) = & \frac{\sqrt{\bar{\alpha}_t - 1}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t
\end{aligned}
$$

И по анологии для $q(f_{t-1}|f_t, f_0)$

$$
\begin{aligned}
q(f_{t-1}|f_t, f_0) &= \mathcal{N}(\tilde{\mu}_t(f_t, f_0), \tilde{\beta}_t I) \\
\tilde{\mu}_t(f_t, f_0) &= \frac{\sqrt{\bar{\alpha}_t - 1}\beta_t}{1 - \bar{\alpha}_t} f_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} f_t
\end{aligned}
$$

$\beta_t$ позволяет выбирать скорость зашумления исходного изображения и коэффициенты $\alpha_t, \bar{\alpha}_t, \tilde{\beta}_t$ определенны как

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t, \quad \alpha_t = 1-\beta_t, \quad \bar{\alpha}_t = \prod_{s=1}^{t}\alpha_t$$

## Обратный процесс

Следуя так же [1]

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(\mu_\theta(x_t,t), \Sigma_\theta(x_t,t))$$

$$p_\theta(f_{t-1}|x_t) = \mathcal{N}(f(\mu_\theta(x_t,t)), \Sigma_\theta(x_t,t))$$

Пусть $p(f_t|x_t)(x) = \delta(x - f(x_t))$, где

$$\delta(0) = 1, \delta(x) = 0, x \neq 0$$

Тогда распределение $p(f_{t-1}|f_t)$ записывается в виде

$$p(f_{t-1}|f_t) = \int p(f_{t-1}|x_{t-1})p(x_{t-1}|x_t)dx_t = p(f_{t-1}|x_t = f^{-1}(f_t))$$

Пусть $\Sigma_\theta = \sigma_t^2 I$ - необучаемая матрица. Среднее значение $\mu_\theta$ параметризуется нейронной сетью, которая оптимизирует вариацонную оценку.

$$\begin{aligned}
L_{t-1} &= L_{t-1}^x + L_{t-1}^f + C \\
L_{t-1}^x &= \mathbb{E}_q\left[\frac{1}{2\sigma_t^2}||\tilde{\mu}_t(x_t,x_0) - \mu_\theta(x_t,t)||^2\right] \\
L_{t-1}^f &= \mathbb{E}_q\left[\frac{1}{2\sigma_t^2}||\tilde{\mu}_t(f_t,f_0) - f(\mu_\theta(x_t,t))||^2\right]
\end{aligned}$$

Рассмотрим параметризацию из [1]

$$\mu_\theta(x_t,t) = \tilde{\mu}_t\left(x_t, \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t))\right) = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t)\right)$$

$$\begin{aligned}
f(\mu_\theta(x_t,t)) &= f\left(\frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t)\right)\right) = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(f(x_t) - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}f(\epsilon_\theta(x_t))\right) \\
&= \tilde{\mu}_t\left(f(x_t), \frac{1}{\sqrt{\bar{\alpha}_t}}(f(x_t) - \sqrt{1-\bar{\alpha}_t}f(\epsilon_\theta(x_t)))\right)
\end{aligned}$$

Тогда для $L_{t-1}^f$ имеем

$$\begin{aligned}
L_{t-1}^f =&\mathbb{E}_{f_t,\epsilon,x_t}\left[\frac{1}{2\sigma_t^2}||\tilde{\mu}_t(f_t,\frac{1}{\sqrt{\bar{\alpha}_t}}(f_t-\sqrt{1-\bar{\alpha}_t}\epsilon))\right.\\
&\left.-\tilde{\mu}_t\left(f(x_t),\frac{1}{\sqrt{\bar{\alpha}_t}}(f(x_t)-\sqrt{1-\bar{\alpha}_t}f(\epsilon_\theta(x_t)))\right)||^2\right]\\
=&\mathbb{E}_{f_t,\epsilon,x_t}\left[\frac{1}{2\sigma_t^2}||\frac{1}{\sqrt{\bar{\alpha}_t}}(f_t-f(x_t))-\frac{\beta_t}{\sqrt{\alpha_t(1-\bar{\alpha}_t)}}(\epsilon-f(\epsilon_\theta(x_t)))||^2\right]
\end{aligned}$$

Так как $x_t = f^{-1}(f_t)$

$$f_t =\sqrt{\bar{\alpha}_t}f_0+\sqrt{1-\bar{\alpha}_t}\epsilon_f$$

$$\begin{aligned}
x_t =&f^{-1}(\sqrt{\bar{\alpha}_t}f_0+\sqrt{1-\bar{\alpha}_t}\epsilon_f)\\
=&\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}f^{-1}(\epsilon_f)
\end{aligned}$$

$$\begin{aligned}
L_{t-1}^f =&\mathbb{E}_{f_0,\epsilon_f,x_0,\epsilon_x}\left[\frac{1}{2\sigma_t^2}||\frac{1}{\sqrt{\bar{\alpha}_t}}(\sqrt{\bar{\alpha}_t}f_0+\sqrt{1-\bar{\alpha}_t}\epsilon_f-f(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_x))\right.\\
&\left.-\frac{\beta_t}{\sqrt{\alpha_t(1-\bar{\alpha}_t)}}(\epsilon_f-f(\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_x)))||^2\right]\\
=&\mathbb{E}_{f_0,\epsilon_f,x_0}\left[\frac{1}{2\sigma_t^2}||\frac{1}{\sqrt{\bar{\alpha}_t}}(\sqrt{\bar{\alpha}_t}f_0+\sqrt{1-\bar{\alpha}_t}\epsilon_f-f(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_x))\right.\\
&\left.-\frac{\beta_t}{\sqrt{\alpha_t(1-\bar{\alpha}_t)}}(\epsilon_f-f(\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_x)))||^2\right]\\
=&\mathbb{E}_{f_0,\epsilon_f,x_0,\epsilon_x}\left[\frac{1}{2\sigma_t^2}||\frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}(\epsilon_f-f(\epsilon_x))\right.\\
&\left.-\frac{\beta_t}{\sqrt{\alpha_t(1-\bar{\alpha}_t)}}(\epsilon_f-f(\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_x)))||^2\right]\\
=&\mathbb{E}_{f_0,\epsilon_f,x_0,\epsilon_x}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}||\epsilon_f-f\left(\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}f^{-1}(\epsilon_f))\right)||^2\right]
\end{aligned}$$

Наконец для оптимизации совместной диффузии $(x, f(x))$, где $f(x)$ линейная бекция и $f(\mathcal{N}(0,I)) = \mathcal{N}(0,I)$ мы имеем функцию потерь

$$L_{t-1} = L_x + L_f \qquad (2)$$

$$\begin{aligned}
L_x &= \mathbb{E}_{x_0,\epsilon_x}\left[||\epsilon_x-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon_x)||^2\right]\\
L_f &= \mathbb{E}_{x_0,\epsilon_f}\left[||\epsilon_f-f\left(\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}f^{-1}(\epsilon_f))\right)||^2\right]
\end{aligned}$$

**Algorithm 1** Train

repeat
  $x_0 \sim q(x_0)$
  $t \sim \text{Uniform}(1, .., T)$
  $\epsilon \sim \mathcal{N}(0, I)$
  $\epsilon_f \sim \mathcal{N}(0, I)$
  $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon_x$
  $x_t^f = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} f^{-1}(\epsilon_f)...$
  Gradient step

$$\nabla ||\epsilon - \epsilon_\theta(x_t)||^2 + ||\epsilon_f - f(\epsilon_\theta(x_t^f))||^2$$

until converged

**Algorithm 2** Sample

$x_T \sim \mathcal{N}(0, I)$
**for** $t = T, .., 1$ **do**
  $z \sim \mathcal{N}(0, I)$ if $t > 1$ else $z = 0$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t)\right) + \sigma_t z$$

**end for**

## Преобразование Фурье

Для наших целей хотелось бы приблизить спектр что бы получить лучшее решение задачи супер резолюции. Быстрое преобразование фурье можно выразить в виде линейного оператора. Матрица характерезующая DFT (Discrete Fourier Transform), однако, не является ортогональной. Более того преобразование фурье отображает пространство комплексных чисел само в себя. То есть, если $\mathcal{F} : \mathbb{C}^n \to \mathbb{C}^n$ это дискретное преобразование фурье, то

$$\mathcal{F}(\epsilon) \sim \mathcal{CN}(0, nI), \epsilon \sim \mathcal{CN}(0, I)$$

Однако в нашем случае $\mathcal{F} : \mathbb{R}^n \to \mathbb{C}$. Это накладывает довольно сильные ограничения на распределение $p_{\epsilon_f}$

$$p_{\epsilon_f} = \mathcal{F}(\epsilon), \epsilon \sim \mathcal{N}(0, 1)$$

Можно заметить, например, что первый элемент вектора ( нумерация начинается с нуля ) не является комплексным числом $\mathcal{F}(\epsilon)_0 = \sum_{i=0}^n \epsilon_i \in \mathbb{R}$, так как в матрице DFT первая строка состоит из 1. Более того, элементы вектора $\mathcal{F}(\epsilon)_i = \overline{\mathcal{F}(\epsilon)_{n-i}}, 0 < i \leqslant \lfloor \frac{n}{2} \rfloor$. Где $\overline{*}$ - это комплексное сопряжение.

Что бы определить диффузию в пространстве спектра я предлагаю рассмотреть следующий оператор $\mathcal{F}_R : \mathbb{R}^n \to \mathbb{R}^n$

$$\mathcal{F}_R(x) = SFx$$

$$S = \sqrt{\frac{2}{N}} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & I_{\lfloor \frac{n}{2} \rfloor} \end{pmatrix}$$

В случае, когда, $\exists k | n = 2k+1$,

$$F = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \cos(-2\pi \frac{1}{n}) & \cos\left(-2\pi\frac{2}{n}\right) & \dots & \cos\left(-2\pi\frac{n-1}{n}\right) \\ 0 & \sin(-2\pi \frac{1}{n}) & \sin\left(-2\pi\frac{2}{n}\right) & \dots & \sin\left(-2\pi\frac{n-1}{n}\right) \\ 1 & \cos(-4\pi \frac{1}{n}) & \cos\left(-4\pi\frac{2}{n}\right) & \dots & \cos\left(-4\pi\frac{n-1}{n}\right) \\ 0 & \sin(-4\pi \frac{1}{n}) & \sin\left(-4\pi\frac{2}{n}\right) & \dots & \sin\left(-4\pi\frac{n-1}{n}\right) \\ & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{1}{n}\right) & \cos\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{2}{n}\right) & \dots & \cos\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{n-1}{n}\right) \\ 0 & \sin\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{1}{n}\right) & \sin\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{2}{n}\right) & \dots & \sin\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{n-1}{n}\right) \end{pmatrix}$$

Если же, $\exists k | n = 2k$,

$$F = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \cos(-2\pi \frac{1}{n}) & \cos\left(-2\pi\frac{2}{n}\right) & \dots & \cos\left(-2\pi\frac{n-1}{n}\right) \\ 0 & \sin(-2\pi \frac{1}{n}) & \sin\left(-2\pi\frac{2}{n}\right) & \dots & \sin\left(-2\pi\frac{n-1}{n}\right) \\ 1 & \cos(-4\pi \frac{1}{n}) & \cos\left(-4\pi\frac{2}{n}\right) & \dots & \cos\left(-4\pi\frac{n-1}{n}\right) \\ 0 & \sin(-4\pi \frac{1}{n}) & \sin\left(-4\pi\frac{2}{n}\right) & \dots & \sin\left(-4\pi\frac{n-1}{n}\right) \\ & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{1}{n}\right) & \cos\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{2}{n}\right) & \dots & \cos\left(-2\pi\lfloor\frac{n}{2}\rfloor\frac{n-1}{n}\right) \end{pmatrix}$$

Этот оператор отображает вещественный вектор в первые $1 + \lfloor\frac{n}{2}\rfloor$ координат его спектра. Так как все координаты после линейно зависят от первых $1 + \lfloor\frac{n}{2}\rfloor$ такое отображение полностью описывает распределение значений спектра. Заметим, что матрица $SF$ - ортогональная и $\mathcal{F}_R^{-1}(x) = F^T S x$ Из ортогональности матрицы следует, что если $\epsilon \sim N(0, I_n)$, то $\mathcal{F}_R(\epsilon) \sim \mathcal{N}(0, I_n)$. Тогда, в соответствие с описаным выше, для диффузионного процесса с дополнительной информацией о спектре можно выписать следующий алгоритм

**Algorithm 3** Train Fourier Diffusion

repeat
  $x_0 \sim q(x_0)$
  $t \sim \text{Uniform}(1, .., T)$
  $\epsilon_x \sim \mathcal{N}(0, I)$
  $\epsilon_f \sim \mathcal{N}(0, I)$
  $x_t^x = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon_x$
  $x_t^f = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} f^{-1}(\epsilon_f)$
  Gradient step

$$\nabla \left[ ||\epsilon_x - \epsilon_\theta(x_t^x)||^2 + ||\epsilon_f - \mathcal{F}_R(\epsilon_\theta(\mathcal{F}_R^{-1}(x_t^f)))||^2 \right]$$

until converged

В дальнейшем можно попробовать рассмотреть диффузию с амплитудой фурье или фазой, но в этих случаях распределения получаются не гауссиановскими, что может создать проблемы при выводе формул.

## Список литературы


[1] Jonathan Ho, Ajay Jain и Pieter Abbeel. «Denoising Diffusion Probabilistic Models». В: *arXiv preprint arxiv:2006.11239* (2020).

[2] Chitwan Saharia и др. «Image super-resolution via iterative refinement». В: *arXiv:2104.07636* (2021).

[3] Alok Pratap Singh. «Numerical Weather Model Super-Resolution». В: 2019. URL: https://api.semanticscholar.org/CorpusID:209334006.

[4] Jascha Sohl-Dickstein и др. «Deep Unsupervised Learning using Nonequilibrium Thermodynamics». В: *Proceedings of the 32nd International Conference on Machine Learning*. Под ред. Francis Bach и David Blei. Т. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, июль 2015, с. 2256—2265. URL: https://proceedings.mlr.press/v37/sohl-dickstein15.html.