

Порождающие модели для прогнозирования наборов временных рядов

Карпев Глеб Андреевич

Московский физико-технический институт

Научный руководитель: д.ф-м.н. В. В. Стрижов

2025

Порождающие модели для прогнозирования наборов временных рядов

Предлагается метод прогнозирования временных рядов с высокой ковариацией и дисперсией

Задача:

Выбрать оптимальную модель для прогнозирования матрицы ковариаций между временными рядами

Предлагаемое решение:

- 1 Строится пространство матриц ковариаций. Метрика удовлетворяет условию Мерсера.
- 2 Выполняется прогноз матрицы ковариаций.
- 3 Восстановление набора временных рядов из пространства матриц ковариаций.

Теоретическая основа генеративного моделирования

- 1 **Римановы порождающие модели** Построение диффузионных процессов на многообразиях SPD-матриц. *Riemannian Score-Based Generative Modelling*
V. De Bortoli et al., NeurIPS 2022
- 2 **Score-based подход** Источник для базового метода score-based моделей. *Generative Modeling by Estimating Gradients*
Y. Song, S. Ermon, NeurIPS 2019
- 3 **Обратная диффузия** Используется для вывода уравнения обратной диффузии в римановых многообразиях. *Time Reversal of Diffusion Processes*
P. Cattiaux et al.

Задача прогнозирования набора временных рядов

Ставится задача прогнозирования набора временных рядов $\mathbf{x}_t = [x_{1t}, \dots, x_{nt}]^T$ с высокой ковариацией и дисперсией. В каждый момент времени набор временных рядов ставится в соответствие матрица ковариаций

$$\mathbf{C}_t = \frac{1}{t-1} \mathbf{Y}_t \mathbf{Y}_t^T,$$

где $\mathbf{Y}_t = [\mathbf{x}_1 - \mu_t, \dots, \mathbf{x}_t - \mu_t]$, $\mu_t = \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i$.
Найти отображение в пространстве $\text{Sym}^+(n)$:

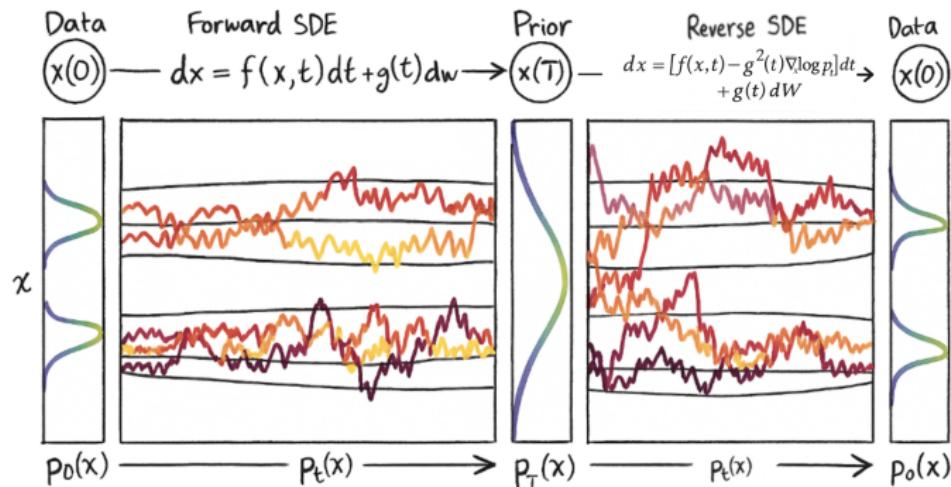
$$f : \underbrace{\mathbf{C}_t}_{\text{Текущая ковариация}} \mapsto \underbrace{\mathbf{C}_{t+\Delta}}_{\text{Прогноз}}$$

с минимальной геодезической ошибкой:

$$\mathcal{L} = \mathbb{E} [d_{Riem}(\mathbf{C}_{t+\Delta}^{true}, f(\mathbf{C}_t))]$$

Порождающие модели для матриц ковариаций

Score-based модель: Итеративный метод генерации *SPD*-матриц через стохастическую диффузию (зашумление) и денойзинг с оценкой градиента плотности.



Риманова геометрия пространства ковариационных матриц

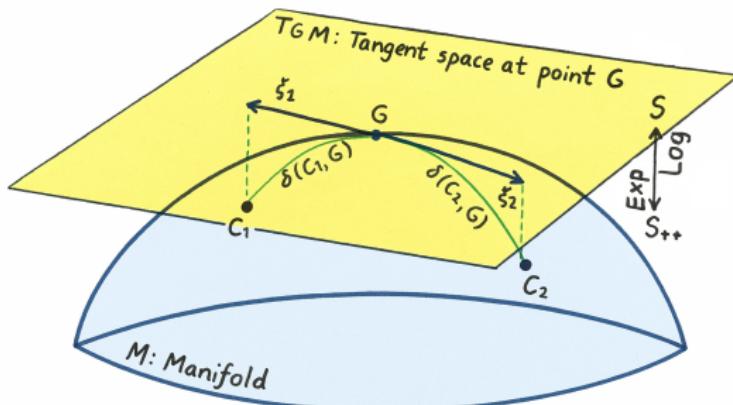
Метрика и расстояния в $\text{Sym}^+(n)$

Аффинно-инвариантная метрика: $\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbf{C}} = \text{tr}(\mathbf{C}^{-1} \mathbf{A} \mathbf{C}^{-1} \mathbf{B})$

Геодезическое расстояние: $d(\mathbf{C}_1, \mathbf{C}_2) = \|\log(\mathbf{C}_1^{-1/2} \mathbf{C}_2 \mathbf{C}_1^{-1/2})\|_F$

Логарифмическое отображение

$$\text{Log}_{\mathbf{G}}(\mathbf{C}) = \mathbf{C}^{1/2} \log(\mathbf{C}^{-1/2} \mathbf{G} \mathbf{C}^{-1/2}) \mathbf{C}^{1/2}$$



Римановы порождающие модели

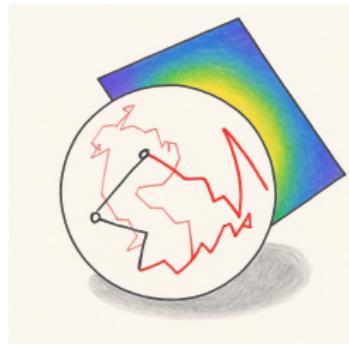
Ключевое улучшение: Учет геометрии SPD-многообразия и сохранение положительной определённости.

Прямой процесс на многообразии

$$d\mathbf{C}_t = b(\mathbf{C}_t)dt + g(t) \underbrace{d\mathbf{B}_t^{\mathcal{M}}}_{\text{Риманов шум}}$$

Обратный процесс на многообразии

$$d\mathbf{C}_t = \left[-b(\mathbf{C}_t) + g(t)^2 \underbrace{\nabla_{\mathcal{M}} \log p_t(\mathbf{C}_t)}_{\text{Риманова score-функция}} \right] dt + g(t)d\bar{\mathbf{B}}_t^{\mathcal{M}}$$



Условное прогнозирование матриц ковариаций

Метод семплирования

1. Инициализация гауссовским шумом:

$$\mathbf{C}_{t+\Delta}^{(0)} \sim \mathcal{N}(\mathbf{0}, \sigma^2(\Delta)\mathbf{I}) \cap \text{Sym}^+(n)$$

2. Итеративное уточнение через обратный SDE:

$$\mathbf{C}_{t+\Delta}^{(k+1)} = \text{Solver}\left(s_\theta(\cdot, \tau_k, \mathbf{C}_t), \mathbf{C}_{t+\Delta}^{(k)}, \tau_k\right), \quad k = 0, \dots, K-1$$

3. Проекция на многообразие SPD (для SGM):

$$\mathbf{C}_{t+\Delta}^{gen} = \text{Proj}_{\text{Sym}^+} \left(\mathbf{C}_{t+\Delta}^{(K)} \right)$$

Вычислительный эксперимент [Синтетические данные]

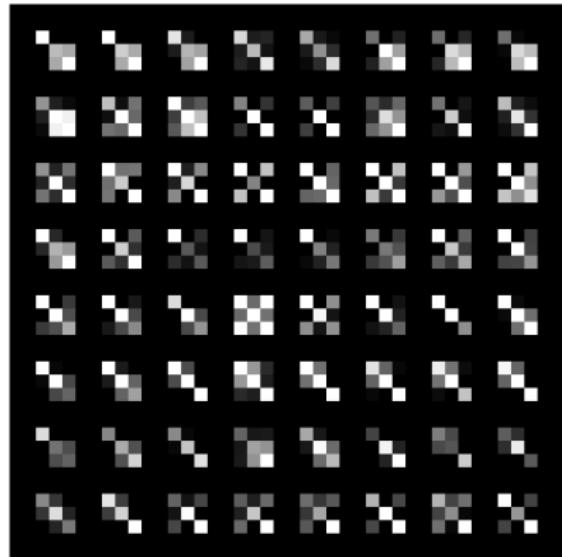


Рис.: Матрицы ковариаций, сэмплированные RSGM

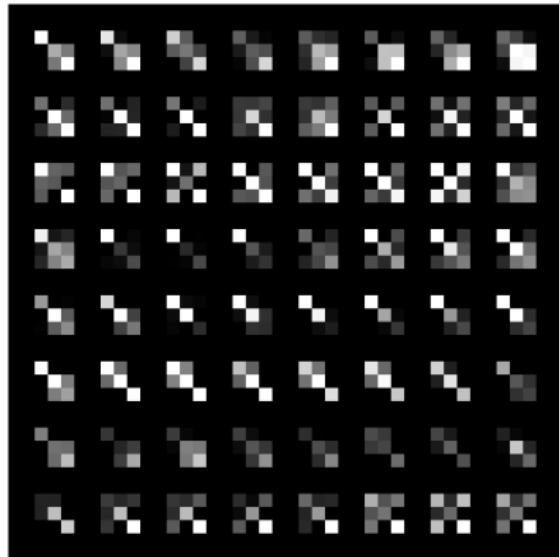


Рис.: Реальные матрицы ковариаций

Наивная модель	SGM	RSGM
1.191	1.376	1.083

$$\mathbb{E} [d_{\text{Riem}} (\mathbf{C}_{t+\Delta}^{\text{true}}, f(\mathbf{C}_t))]$$

Выносится на защиту

- Предложен метод, который выполняет кодирование временных рядов с помощью матрицы ковариаций, выполняет прогноз, а затем выполняет декодирование полученной матрицы.
- Ошибка предложенного метода генерации матрицы попарных ковариаций:

	Наивная модель	SGM	RSGM
$\mathbb{E} [d_{\text{Riem}} (\mathbf{C}_{t+\Delta}^{\text{true}}, f(\mathbf{C}_t))]$	1.191	1.376	1.083