УДК 519.65; 519.613; 519.246 doi 10.26089/NumMet.v18r319

ПРИЛОЖЕНИЕ БЛОЧНО-МАЛОРАНГОВЫХ МАТРИЦ ДЛЯ ЗАДАЧИ РЕГРЕССИИ НА ОСНОВЕ ГАУССОВСКИХ ПРОЦЕССОВ

Д. А. Сушникова[1]

Рассматривается задача регрессии на основе гауссовских процессов. В ходе моделирования коррелированного шума при помощи гауссовского процесса основной проблемой является подсчет апостериорного среднего и дисперсии прогноза, для чего необходимо обращать плотную матрицу ковариации размера $n \times n$, где n — размер выборки. Кроме того, для оценки правдоподобия требуется вычислять логарифм определителя плотной ковариационной матрицы, что тоже является трудоемкой задачей. Предложен метод быстрого вычисления логарифма определителя матрицы ковариации на основе идеи ее аппроксимации разреженной матрицей. При сравнении с методом HODLR (Hierarchically Off-Diagonal Low-Rank) и с наивным традиционным методом предложенный метод оказался более эффективным по времени.

Ключевые слова: гауссовские процессы, $\mathcal{H}^2$-матрица, разреженная матрица, разложение Холецкого.

1. Введение. Одной из актуальных задач прикладной статистики является задача регрессии, в которой набор данных имеет неизвестные корреляции в шуме [3]. Эффект этого коррелированного шума часто трудно оценить, однако его влияние на окончательный ответ может быть велико. В настоящей статье рассматривается искусственно сгенерированный набор данных с коррелированным шумом и простой нелинейной моделью. Ковариационная структура в данных моделируется с использованием гауссовского процесса.

В ходе применения гауссовского процесса к задаче регрессии требуется многократно вычислять значение функции правдоподобия для различных наборов параметров, а следовательно, необходимо вычислять логарифм определителя матрицы ковариации и решать с ней линейные системы. Матрица ковариации, вообще говоря, является плотной, поэтому как вычисление логарифма ее определителя, так и решение системы с ней является трудоемкой задачей. Существует несколько методов аппроксимации регрессии на основе гауссовских процессов в машинном обучении [9, 10]; в работах [11, 12], например, делается аппроксимация детерминанта с помощью методов Монте-Карло. В работе [2] показано, что матрица ковариации обладает $\mathcal{H}^2$-структурой. В нашей работе предлагается аппроксимировать эту матрицу в $\mathcal{H}^2$-формате, затем приводить ее к разреженному виду при помощи метода, описанного в работе [4], а потом приближенно факторизовать при помощи пакета CHOLMOD [5] и затем считать логарифм ее определителя и решать с ней систему. В работе [2] определитель матрицы ковариации предлагалось считать с помощью аппроксимации в формате HODLR (Hierarchically Off-Diagonal Low-Rank) [1, 2] и последующей ее факторизации. В настоящей статье будет показано, что предложенный автором метод подсчета определителя является более выгодным по времени и по памяти.

Приведем определение гауссовского процесса.

Определение 1. Гауссовский процесс (ГП) — это совокупность случайных величин, любое конечное число которых имеет совместное гауссово распределение. Случайные величины из ГП должны обладать свойством согласованности. Это означает, что если ГП задает $y^{(1)}, y^{(2)} \sim \mathcal{N}(\mu, \Sigma)$, то также задается $y^{(1)} \sim \mathcal{N}(\mu_1, \Sigma_{11})$, где $\mu = \begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}$, $\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$. ГП полностью определяется средней функцией и положительно определенной матрицей ковариации.

Подробное описание процесса и его приложение в задаче регрессии можно найти в работе [3].

2. Простой одномерный пример. Для моделирования коррелированного шума при помощи гауссовского процесса, а также для решения задачи регрессии в нашей работе используется пакет George [6]. Все вычисления проведены в однопроцессорном режиме на сервере с 32 Intel® Xeon® E5-2640 v2 (20M Cache, 2.00 GHz) процессорами и с 256GB RAM.

[1] Сколковский институт науки и технологий, ул. Нобеля, 3, 143026, Москва; стажер-исследователь, e-mail: d.sushnikova@skoltech.ru

Наборы данных генерируются следующим образом: для N случайных точек t_i, $i \in \overline{1,\dots,N}$, генерируются значения функции $y_i = f(t_i)$, где $f = \alpha \exp\left(-\dfrac{[t-l]^2}{m}\right)$, $\alpha = -1$ — амплитуда, $l = 0.1$ — смещение, $m = 0.4$ — ширина. Шум генерируется при помощи гауссовского процесса с матрицей ковариации $K_{ij} = \sigma_i^2 \delta_{ij} + k(r)$, где $\sigma_i^2 = 0.3$. Ядро $k(r)$ задается в виде $k(r_{ij}) = \exp\left(-\dfrac{r^2}{2}\right)$, где $r_{ij} = |t_i - t_j|$.

Для полученных искусственных данных ищется функция в следующем виде:

$$f_\theta = ct + d + \alpha \exp\left(-\frac{[t-l]^2}{m}\right).$$

В нашем примере (рис. 1) шум предполагается коррелированным и моделируется при помощи гауссовского процесса с матрицей ковариации

$$K_{ij} = \sigma_i^2 \delta_{ij} + k(r) \tag{1}$$

с ядром

$$k(r_{ij}) = b^2 \left(1 + \frac{\sqrt{3}r}{\tau}\right) \exp\left(-\frac{\sqrt{3}r}{\tau}\right), \tag{2}$$

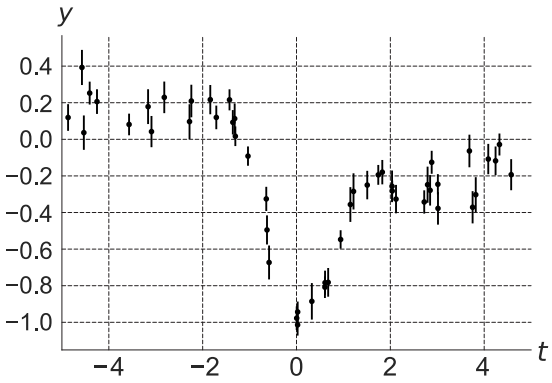

Рис. 1. Пример набора данных для указанных выше параметров, $N = 50$

где $r_{ij} = |t_i - t_j|$, b^2 и τ — параметры модели. Таким образом, по исходным данным t_i, y_i, $i \in \overline{1,\dots,N}$, требуется восстановить вектор параметров $\theta = \{c, d, \alpha, l, m, b, \tau, \sigma_i\}$.

Запишем функцию правдоподобия

$$\ln p\Big(\{y_i\}\big|\{t_i\}, \{\sigma_i^2\}, \theta\Big) = -\frac{1}{2}\, e^\top K^{-1} e - \frac{1}{2} \ln\det K - \frac{N}{2} \ln 2\pi, \tag{3}$$

где

$$e = \begin{bmatrix} y_1 - f_\theta(t_1) \\ y_2 - f_\theta(t_2) \\ \vdots \\ y_N - f_\theta(t_N) \end{bmatrix}. \tag{4}$$

Параметры определяются при помощи метода Монте-Карло марковских цепей (Markov Chain Monte Carlo, MCMC) [7].

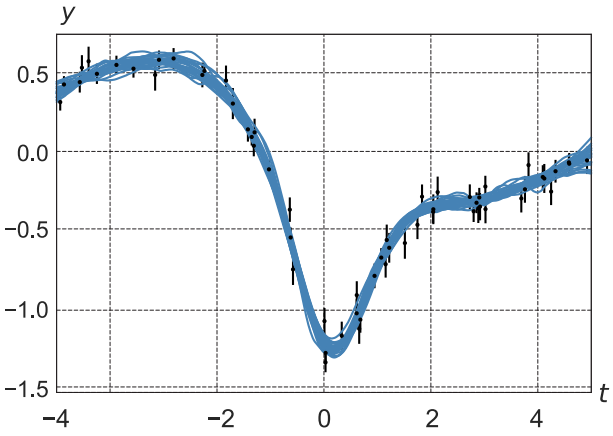

Рис. 2. Результаты регрессии для коррелированного шума

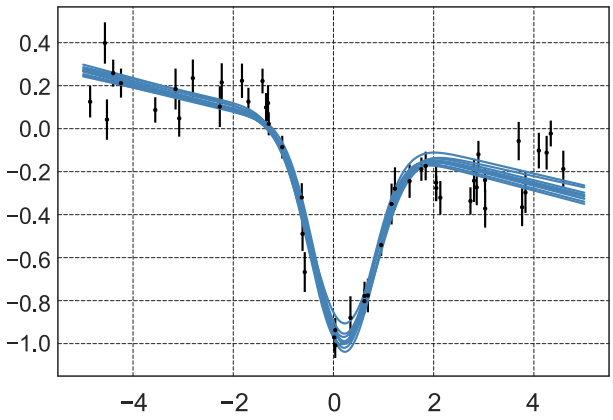

Рис. 3. Результаты регрессии для некоррелированного шума

На рис. 2 приведены функции, сгенерированные при помощи 24 случайно выбранных наборов оцененных параметров. Для сравнения также рассмотрим случай, когда шум предполагается некоррелированным, тогда функция правдоподобия будет иметь следующий вид:

$$\ln p\Big(\{y_i\}\big|\{t_i\}, \{\sigma_i^2\}, \theta\Big) = -\frac{1}{2} \sum_{i=1}^{N} \frac{\big(y_i - f_\theta(t_i)\big)^2}{\sigma_i^2}.$$

Параметры в методе с некоррелированным шумом определяются тоже при помощи метода Монте-Карло марковских цепей.

На рис. 3 приведены функции, сгенерированные при помощи 24 случайно выбранных наборов оцененных параметров для задачи, в которой шум предполагался некоррелированным. Метод, при котором шум предполагался некоррелированным, уступает в точности методу с коррелированным шумом. Рассмотрим параметры, полученные в ходе регрессии двух этих случаях.

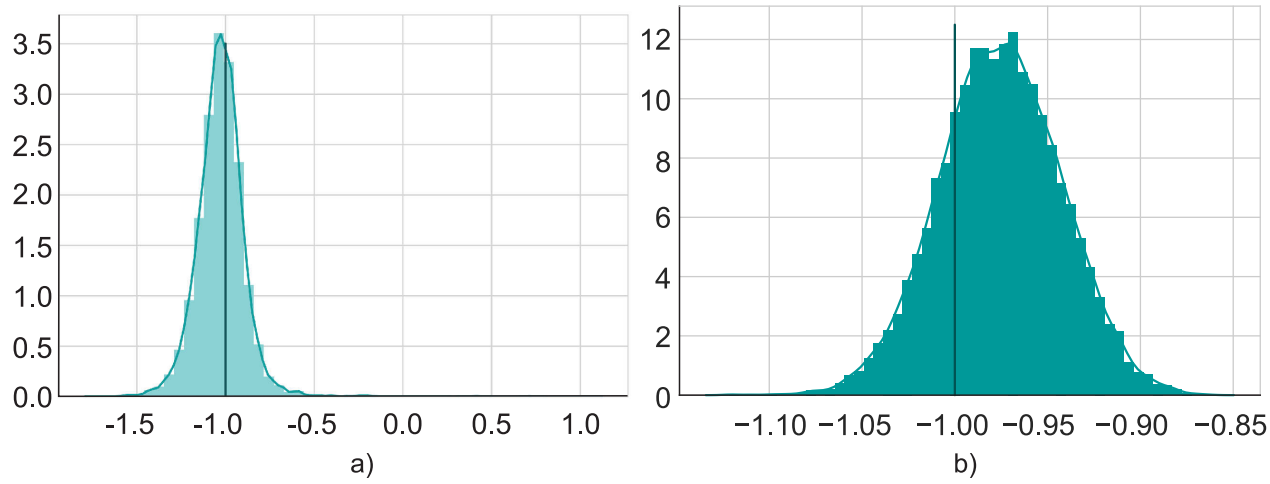

Рис. 4. Распределение параметра α для разных типов моделирования шума:
a) коррелированный шум, b) некоррелированный шум

На рис. 4 прямая линия отмечает верное значение параметра α. Для некоррелированного шума параметр оценен с большой погрешностью. Рассмотрим два других параметра l и σ.

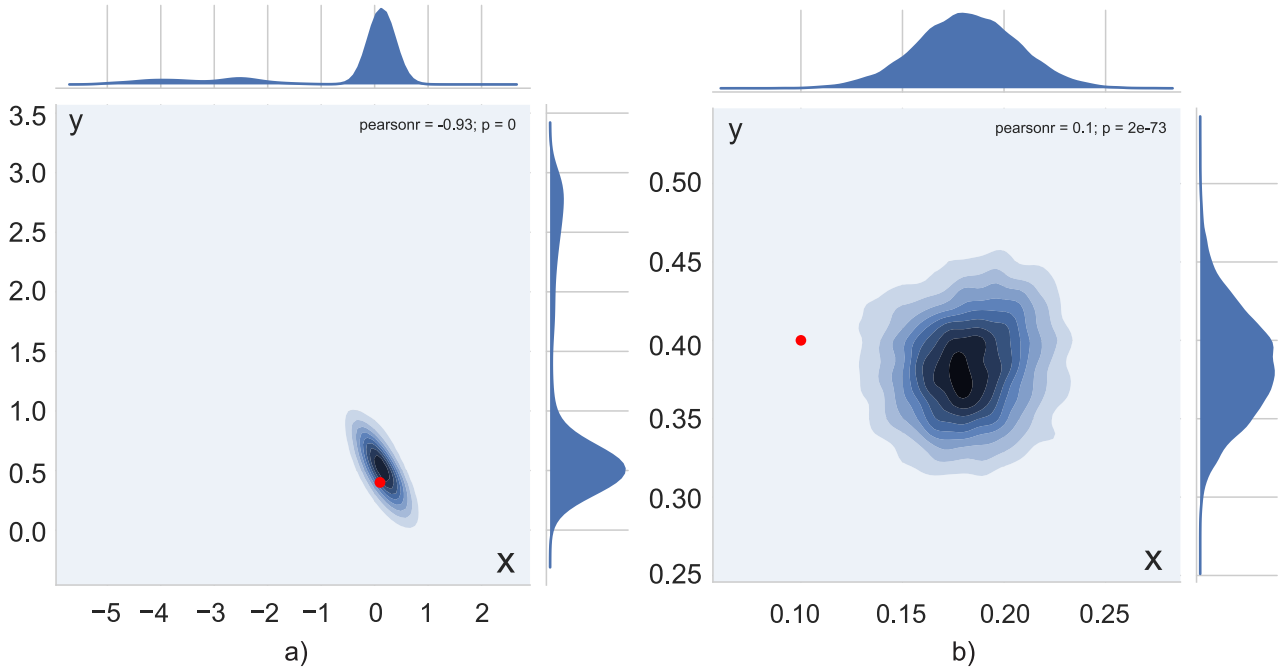

Рис. 5. Распределение параметров l и σ для разных типов моделирования шума:
a) коррелированный шум, b) некоррелированный шум

На рис. 5 красная точка отмечает верное значение параметров l и σ. Для некоррелированного шума параметры оценены с большой погрешностью. Графики, приведенные выше, показывают, что метод с моделированием коррелированного шума (т.е. моделирование шума при помощи гауссовского процесса) более эффективен для оценки параметров, чем метод, предполагающий белый шум. Оценим коэффициент детерминации двух рассмотренных выше моделей.

Коэффициент детерминации R^2 оценивает, с какой вероятностью модель будет хорошо предсказывать будущие результаты измерений. Наилучший возможный результат — это 1.0, коэффициент детерминации R^2 может быть и отрицательным. Константная модель, которая всегда предсказывает ожидаемое значение y вне зависимости от входных параметров, будет иметь $R^2 = 0.0$.

Определение 2 (коэффициент детерминации). Если $\hat{y}_i$ — предсказанное значение i-х результатов измерений и y_i — соответствующее верное значение, то тогда параметр R^2, оцененный на n_s наборах измерений, определяется как

$$R^2(y, \hat{y}) = 1 - \left(\sum_{i=0}^{n_s-1} (y_i - \hat{y}_i)^2\right)\left(\sum_{i=0}^{n_s-1} (y_i - \bar{y}_i)^2\right)^{-1}, \quad \text{где} \quad \bar{y}_i = \frac{1}{n_s}\sum_{i=0}^{n_s-1} y_i.$$

В табл. 1 для разного числа измерений n_s приведены коэффициенты детерминации для моделей, рассмотренных выше.

Таблица 1

Коэффициент детерминации для моделей с коррелированным шумом и без него

n_s	50	100	200	400
Модель с кор. шумом	$1 - 3.4 \times 10^{-6}$	-2.8×10^{-5}	$1 - 1.2 \times 10^{-5}$	$1 - 6.3 \times 10^{-6}$
Модель с некор. шумом	$1 - 7.3 \times 10^{-2}$	$1 - 7.2 \times 10^{-2}$	$1 - 9.1 \times 10^{-2}$	$1 - 7.5 \times 10^{-2}$

Таким образом, модель с коррелированным шумом более эффективна для оценки параметров. Далее для 2D- и 3D-задач рассматривается только данная модель. Отметим, что логарифм определителя матрицы ковариации и решение системы с матрицей ковариации требуется вычислять именно в этой модели.

3. Двумерная задача. Для двумерной задачи регрессии наборы данных генерируются аналогично одномерному случаю следующим образом: для N случайных точек $t_i \in \mathbb{R}^2$, $i \in \overline{1, \ldots, N}$, генерируются значения функции $y_i = f(t_i)$, где $f = \alpha \exp\left(-\dfrac{\left[\|t\| - l\right]^2}{m}\right)$, $\alpha = -1$ — амплитуда, $l = 0.1$ — смещение, $m = 0.4$ — ширина. Данные зашумляются, как и в одномерном случае, при помощи гауссовского процесса с матрицей ковариации $K_{ij} = \sigma_i^2 \delta_{ij} + k(r)$, где $\sigma_i^2 = 0.3$. Ядро $k(r)$ задается в виде $k(r_{ij}) = \exp\left(-\dfrac{r^2}{2}\right)$, где $r_{ij} = \|t_i - t_j\|$.

Для полученных синтетических данных решается задача регрессии, а функция ищется в виде

$$f_\theta = c \cdot t + d + \alpha \exp\left(-\frac{\left[\|t\| - l\right]^2}{m}\right), \quad \text{где} \quad c \in \mathbb{R}^2, \quad d \in \mathbb{R}.$$

Шум предполагается коррелированным и моделируется при помощи гауссовского процесса с матрицей ковариации такой же, как и в одномерной задаче (1) с ядром (2). Таким образом, по исходным данным t_i, y_i, $i \in \overline{1, \ldots, N}$, требуется восстановить вектор параметров $\theta = \{c, d, \alpha, l, m, b, \tau, \sigma_i\}$. Функция правдоподобия имеет вид (3), где e представляется в виде (4). Параметры определяются при помощи метода Монте-Карло марковских цепей. В процессе подсчета требуется считать логарифм определителя плотной матрицы K и решать систему с ней. Так как матрица ковариации симметричная и положительно определенная, для этого необходимо найти ее факторизацию Холецкого. По умолчанию в пакете George матрица факторизуется при помощи пакета NumPy [8] на основе инструментов для плотных матриц, что приводит к значительным затратам по памяти и по времени. Кроме того, в пакет встроен решатель HODLR, который использует блочно-малоранговые техники для быстрой факторизации; однако, поскольку метод HODLR — это аналог одномерных $\mathcal{H}^2$-матриц, для 2D и 3D задач он не эффективен. Мы предлагаем использовать для аппроксимации разреженное представление $\mathcal{H}^2$-матрицы, факторизованное с помощью разреженных инструментов; будем называть этот метод "спарсификация $\mathcal{H}^2$". Сравним время подсчета логарифма определителя матрицы K для методов HODLR и спарсификации $\mathcal{H}^2$для параметра точности аппроксимации $\varepsilon = 10^{-2}$ с учетом вклада временны́х затрат метода спарсификации $\mathcal{H}^2$на разных этапах вычислений: аппроксимация в виде разреженной матрицы, подсчет факторизации и вычисление логарифма определителя (рис. 6).

Отметим, что кроме вычисления логарифма определителя необходимо вычислять и решение системы с матрицей K, однако после построения факторизации обе эти задачи решаются за незначительное время.

Сравним общее время вычисления функции правдоподобия для метода HODLR $\left(\varepsilon = 10^{-2}, 10^{-4}\right)$, метода спарсификации $\mathcal{H}^2\left(\varepsilon = 10^{-2}, 10^{-4}\right)$ и плотного метода (рис. 7).

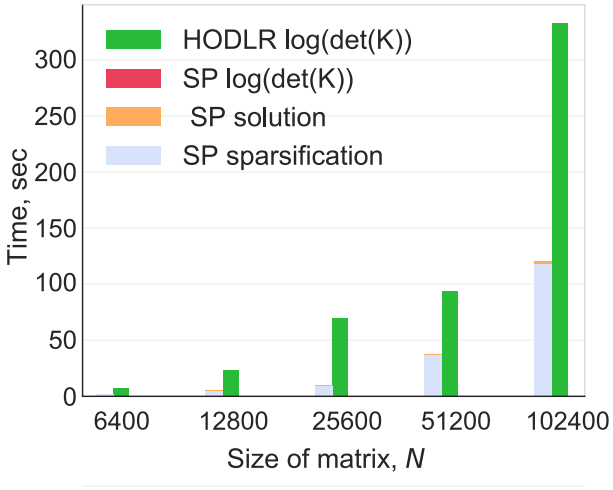

Рис. 6. Время вычисления логарифма определителя матрицы K

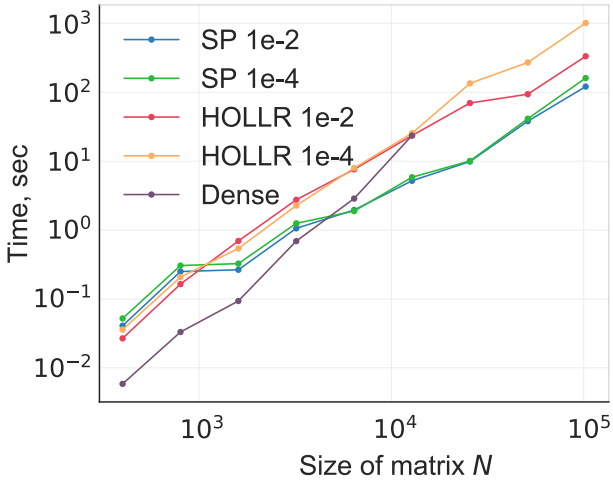

Рис. 7. Время вычисления функции правдоподобия

Отметим, что плотный метод требует слишком много памяти и не работает для матриц размером больше 10^4, однако плотный метод решает системы с хорошей точностью. Покажем, что для данной задачи хорошая точность вычисления функции правдоподобия не требуется. Распределение параметров при замене точного подсчета функции правдоподобия на приближенное изменяется незначительно.

В табл. 2 приведены коэффициенты детерминации для различных способов подсчета функции правдоподобия.

Таблица 2
Коэффициент детерминации для различных методов подсчета правдоподобия

	R^2
HODLR, $\varepsilon = 1e-2$	$1 - 1.8 \times 10^{-3}$
HODLR, $\varepsilon = 1e-4$	$1 - 7.9 \times 10^{-5}$
Сп. $\mathcal{H}^2$, $\varepsilon = 1e-2$	$1 - 8.5 \times 10^{-5}$
Сп. $\mathcal{H}^2$, $\varepsilon = 1e-4$	$1 - 3.4 \times 10^{-5}$
Плотн.	$1 - 8.4 \times 10^{-6}$

Таблица 3
Коэффициент детерминации для модели с коррелированным шумом и без него

	R^2
HODLR, $\varepsilon = 1e-2$	$1 - 2.0 \times 10^{-4}$
HODLR, $\varepsilon = 1e-4$	$1 - 1.2 \times 10^{-4}$
Сп. $\mathcal{H}^2$, $\varepsilon = 1e-2$	$1 - 4.2 \times 10^{-4}$
Сп. $\mathcal{H}^2$, $\varepsilon = 1e-4$	$1 - 4.6 \times 10^{-5}$
Плотн.	$1 - 2.6 \times 10^{-6}$

4. Трехмерная задача. Для трехмерной задачи регрессии наборы данных генерируются аналогично задаче в 2D-случае. Для полученных синтетических данных решается задача регрессии, а функция ищется в следующем виде:

$$f_\theta = c \cdot t + d + \alpha \exp\left(-\frac{\left(\|t\| - l\right)^2}{m}\right),$$

где $c \in \mathbb{R}^3$, $d \in \mathbb{R}$. Шум моделируется при помощи гауссовского процесса матрицей ковариации (1) с ядром (2). Таким образом, по исходным данным t_i, y_i, $i \in \overline{1, \ldots, N}$, требуется восстановить вектор параметров $\theta = \{c, d, \alpha, l, m, b, \tau, \sigma_i\}$. Функция правдоподобия имеет вид (3), где e представляется в виде (4). Параметры определяются при помощи метода Монте-Карло марковских цепей. Для подсчета функции правдоподобия (3), как и в двумерном случае, используются методы HODLR и спарсификация $\mathcal{H}^2$.

Сравним общее время вычисления функции правдоподобия; самый трудоемкий процесс данного вычисления — факторизация плотной матрицы — выполняется при помощи метода HODLR $\left(\varepsilon = 10^{-2}, 10^{-4}\right)$, метода спарсификации $\mathcal{H}^2\left(\varepsilon = 10^{-2}, 10^{-4}\right)$ и плотного метода (рис. 8). Кроме того, сравним время подсчета логарифма определителя матрицы K для методов HODLR и спарсификации $\mathcal{H}^2$для параметра точности аппроксимации $\varepsilon = 10^{-2}$ (рис. 9).

Отметим, что плотный метод требует слишком много памяти и не работает для матриц размером больше 10^4, однако плотный метод факторизует матрицу K с высокой точностью. Покажем, что для

данной задачи высокая точность факторизации не требуется. В табл. 3 приведены коэффициенты детерминации для рассмотренных выше моделей и для различных способов факторизации матрицы K.

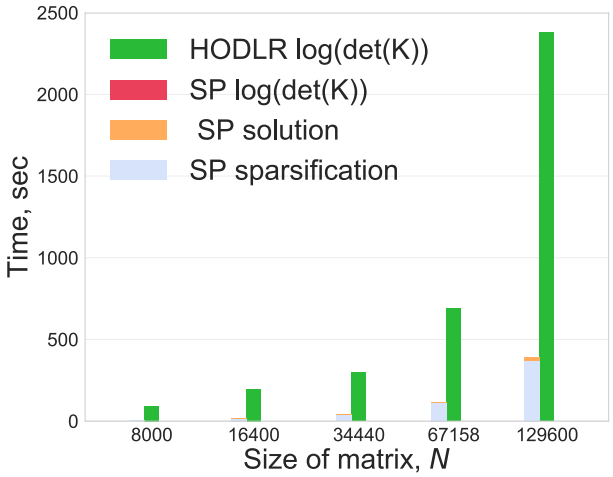

Рис. 8. Время вычисления

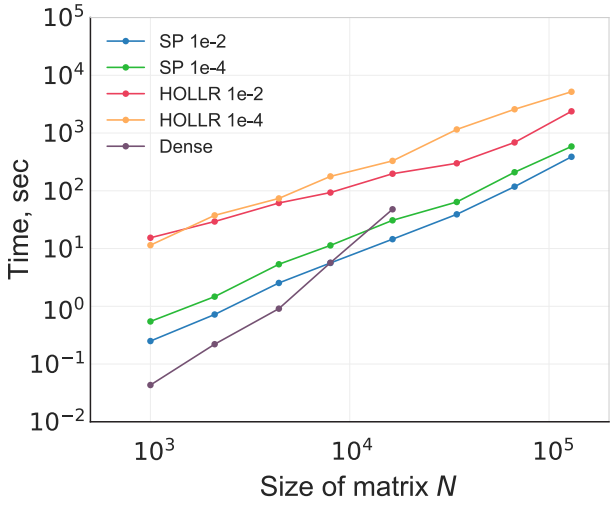

Рис. 9. Время вычисления логарифма определителя матрицы K

5. Заключение. В настоящей статье предложено приложение факторизации иерархических матриц в задаче регрессии на основе гауссовских процессов. Логарифм определителя матрицы ковариации и решение системы с матрицей ковариации вычислялись при помощи аппроксимации в $\mathcal{H}^2$-формате и приближенной факторизации. Тестирование проводилось для двух- и трехмерной задач и для синтетически сгенерированных данных. Проводилось сравнение с методом HODLR. Предложенный метод значительно ускорил процесс вычисления правдоподобия при моделировании коррелированного шума в задаче регрессии.

Разделы 1 и 2 настоящей статьи выполнены при финансовой поддержке РНФ (проект 17–11–01376); разделы 3 и 4 выполнены при финансовой поддержке РФФИ (проект 17–01–00854).

СПИСОК ЛИТЕРАТУРЫ

1. *Ambikasaran S., Darve E.* An $\mathcal{O}(N \log N)$ fast direct solver for partial hierarchically semi-separable matrices // J. Sci. Comput. 2013. **57**, N 3. 477–501.
2. *Ambikasaran S., Foreman-Mackey D., Greengard L., et al.* Fast direct methods for Gaussian processes // arXiv preprint arXiv:1403.6015. 2014.
3. *Rasmussen C.E., Williams C.K.I.* Gaussian processes for machine learning // Cambridge: MIT Press, 2006.
4. *Sushnikova D., Oseledets I.* Simple non-extensive sparsification of the hierarchical matrices // arXiv preprint arXiv:1705.04601. 2017.
5. *Davis T.A., Hager W.W.* Dynamic supernodes in sparse Cholesky update/downdate and triangular solves // ACM Trans. Math. Software. 2009. **35**, N 4. 1–23.
6. *Lalgudi H.G., Bilgin A., Marcellin M.W., et al.* Four-dimensional compression of fMRI using JPEG2000 // SPIE Proc. 2005. **5747**. 1028–1037.
7. *Ambikasaran S., Foreman-Mackey D., Greengard L., et al.* Fast direct methods for Gaussian processes and the analysis of NASA Kepler mission data // arXiv preprint arXiv:1403.6015. 2015.
8. NumPy Package. http://http://www.numpy.org. Cited June 7, 2017.
9. *Беляев М., Бурнаев Е., Капушев Е.* Вычислительно эффективный алгоритм построения регрессии на основе гауссовских процессов в случае структурированных выборок // Ж. вычисл. матем. и матем. физ. 2016. **56**, № 4. 507–522.
10. *Snelson E., Ghahramani Z.* Sparse Gaussian processes using pseudo-inputs // Proc. 18th Int. Conf. on Advances in Neural Information Processing Systems. Vol. 18. Cambridge: MIT Press, 2006. 1257–1264.
11. *Zhang Y., Leithead W.E., Leith D.J., Walshe L.* Log-det approximation based on uniformly distributed seeds and its application to Gaussian process regression // Journal of Computational and Applied Mathematics. 2008. **220**, N 1–2. 198–214.

12. *Leithead W.E., Zhang Y., Leith D.J.* Efficient Gaussian process based on BFGS updating and logdet approximation // IFAC Proceedings Volumes. 2005. **38**, N 1. 1305–1310.

Поступила в редакцию
17.05.2017

Application of Block Low-Rank Matrices in Gaussian Processes for Regression

D. A. Sushnikova [1]

[1] *Skolkovo Institute of Science and Technology; ulitsa Nobelya, 3, Skolkovo Innovation Center, Moscow Region, 143025, Russia; Research Intern, e-mail: d.sushnikova@skoltech.ru*

Received May 17, 2017

Abstract: The Gaussian processes for regression are considered. During simulation of correlated noises using the Gaussian processes, the main difficulty is the computation of the posterior mean and dispersion of the prediction. This computation requires the inversion of the dense covariance matrix of order n, where n is the sample size. In addition, for the likelihood evaluation we need to compute the logarithm of the determinant of the dense covariance matrix, which is also a time-consuming problem. A new method for the fast computation of the covariance matrix logarithm is proposed. This method is based on the approximation of this matrix by a sparse matrix. The proposed method appears to be time efficient compared to the HODLR (Hierarchically Off-Diagonal Low-Rank) method and the traditional dense method.

Keywords: Gaussian processes, $\mathcal{H}^2$ matrix, sparse matrix, Cholesky factorization.

References

1. S. Ambikasaran and E. Darve, "An $\mathcal{O}(N \log N)$ Fast Direct Solver for Partial Hierarchically Semi-Separable Matrices," J. Sci. Comput. **57** (3), 477–501 (2013).

2. S. Ambikasaran, D. Foreman-Mackey, L. Greengard, et al., *Fast Direct Methods for Gaussian Processes* arXiv preprint: 1403.6015 [math.NA] (Cornell Univ. Library, Ithaca, 2014),
available at https://arxiv.org/abs/1403.6015/.

3. C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006).

4. D. Sushnikova and I. Oseledets, *Simple Non-Extensive Sparsification of the Hierarchical Matrices* arXiv preprint: 1705.04601v1 [math.NA] (Cornell Univ. Library, Ithaca, 2017),
available at https://arxiv.org/abs/1705.04601/.

5. T. A. Davis and W. W. Hager, "Dynamic Supernodes in Sparse Cholesky Update/Downdate and Triangular Solves," ACM Trans. Math. Software **35** (4), 1–23 (2009).

6. H. G. Lalgudi, A. Bilgin, M. W. Marcellin, et al., "Four-Dimensional Compression of fMRI Using JPEG2000," SPIE Proc. **5747**, 1028–1037 (2005).

7. S. Ambikasaran, D. Foreman-Mackey, L. Greengard, et al., *Fast Direct Methods for Gaussian Processes and the Analysis of NASA Kepler Mission Data* arXiv preprint: 1403.6015v2 [math.NA] (Cornell Univ. Library, Ithaca, 2015), available at https://arxiv.org/abs/1403.6015/.

8. NumPy Package. http://http://www.numpy.org. Cited June 7, 2017.

9. M. Belyaev, E. Burnaev, and E. Kapushev, "Computationally Efficient Algorithm for Gaussian Process Regression in Case of Structured Samples," Zh. Vychisl. Mat. Mat. Fiz. **56** (4), 507–522 (2016) [Comput. Math. Math. Phys. **56** (4), 499–513 (2016)].

10. E. Snelson and Z. Ghahramani, "Sparse Gaussian Processes Using Pseudo-Inputs," in *Proc. 18th Int. Conf. on Advances in Neural Information Processing Systems, Vancouver, Canada, December 05–08, 2005* (MIT Press, Cambridge, 2006), Vol. 18, pp. 1257–1264.

11. Y. Zhang, W. E. Leithead, D. J. Leith, and L. Walshe, "Log-Det Approximation Based on Uniformly Distributed Seeds and its Application to Gaussian Process Regression," J. Comput. Appl. Math. **220** (1–2), 198–214 (2008).

12. W. E. Leithead, Y. Zhang, and D. J. Leith, "Efficient Gaussian Process Based on BFGS Updating and Logdet Approximation," IFAC Proc. Volumes **38** (1), 1305–1310 (2005).