The Dantzig selector for a linear model of diffusion processes

Kou Fujimori
Waseda University

Abstract

In this talk, a linear model of diffusion processes with unknown drift and diagonal diffusion matrices is discussed. We will consider the estimation problems for unknown parameters based on the discrete time observation in high-dimensional and sparse settings. To estimate drift matrices, the Dantzig selector which was proposed by Candés and Tao in 2007 will be applied. We will prove two types of consistency of the Dantzig selector for the drift matrix; one is the consistency in the sense of l_q norm for every $q \in [1, \infty]$ and another is the variable selection consistency. Moreover, we will construct an asymptotically normal estimator for the drift matrix by using the variable selection consistency of the Dantzig selector.

1 Introduction

Let us consider the following model given by the linear stochastic differential equation:

$$X_t = X_0 + \int_0^t \Theta^\top \phi(X_s) ds + \sigma W_t, \tag{1}$$

where $\{W_t\}_{t \geq 0} := \{(W_t^1, \ldots, W_t^p)\}_{t \geq 0}$ is a p-dimensional standard Brownian motion, Θ is a $p \times p$ sparse deterministic matrix, $\sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ is a $p \times p$ diagonal matrix and $\phi(x) = (\phi_1(x_1), \ldots, \phi_p(x_p))^\top$ for $x = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ is a smooth $\mathbb{R}^p$-valued function. We will propose some estimators for the true values (Θ^0, σ^0) of (Θ, σ) based on the observation of $\{X_t\}_{t \geq 0}$ at $n+1$ equidistant time points $0 =: t_0^n < t_1^n < \ldots < t_n^n$, under the high-dimensional and sparse setting, *i.e.*, $p \gg n$ and the number of nonzero components of the true value Θ^0 is relatively small.

To deal with high-dimensional and sparse parameters, various kinds of estimators for regression models have been discussed. One of the most famous estimation methods is the l_1-penalized method called Lasso proposed originally by Tibshirani (1996), which has been studied for regression models with high-dimensional and sparse parameters in various models including the ones of stochastic processes.

On the other hand, a relatively new estimation procedure called the Dantzig selector was proposed for linear regression models by Candés and Tao (2007) as follows.

$$\hat{\beta}_D := \arg\min_{\beta \in \mathcal{C}} \|\beta\|_1, \quad \mathcal{C} := \left\{ \beta \in \mathbb{R}^p : \sup_{1 \le j \le p} |Z^{j\top}(Y - Z\beta)| \le \lambda \right\},$$

where $\lambda \geq 0$ is a tuning parameter. When $\lambda = 0$, the Dantzig selector coincides with the classical estimators such as the LSE in general cases and the MLE in Gaussian noise cases. For $\lambda > 0$, the Dantzig selector searches for the sparsest β within the given distance of the classical estimators. The Dantzig selector has been studied well especially for *i.i.d.* models. For example, Bickel et al. (2009) showed that the Dantzig selector has some properties similar to Lasso estimator for linear regression models in the sense of the consistency. In addition, as well as Lasso, the Dantzig selector has variable selection consistency for some regression models. Fan et al. (2016) showed the varable selection consistency of the Dantzig selector for general single index models by using the irrepresentable conditions which are obtained from the KKT condition of the optimization problem. The Dantzig selector also has a good potential to be applied for other models including the models of stochastic processes. For instance, Antoniadis et al. (2010) applied this method to estimate regression parameter for Cox's proportional hazards model and proved the obtained estimator has the consistency. Fujimori (2017) studied the variable selection consistency of the Dantzig selector for the proportional hazards model and construct asymptotically normal estimators for the regression parameter and the cumulative baseline hazard function. Moreover, it is well-known that the Dantzig selector for linear models has computational advantages since it can be solved by a linear programming, while Lasso demands a convex program.

In this talk, we will apply the Dantzig selector to the linear models of stochastic processes (1) to estimate the drift matrix Θ^0 and prove the consistency in the sense of l_q norm for every $q \in [1, \infty]$ and the variable selection consistency under some appropriate conditions. Moreover, using the variable selection consistency, we will construct a new estimator which has an asymptotic normality. We can prove the consistency of the Dantzig selector by the standard way which is similar to Bickel et al. (2009). However, since dealing with the KKT conditions of the Dantzig selector for our model is more difficult due to the complicated structure of the model than those of *i.i.d.* models, it may be hard to obtain the same results as Fan et al. (2016) concerning the variable selection consistency. Therefore, we will show another type of variable selection consistency by using a thresholding method.

2 Notation

We denote by $\|\cdot\|_q$ the l_q norm of vector for every $q \in [1, \infty]$, *i.e.*, for $v = (v_1, v_2, \ldots, v_p)^\top \in \mathbb{R}^p$, we define:

$$\|v\|_q = \left(\sum_{j=1}^{p} |v_j|^q \right)^{\frac{1}{q}}, \quad q < \infty;$$

$$\|v\|_\infty = \sup_{1 \le j \le p} |v_j|.$$

In addition, for a $m \times n$ matrix A, where $m,\ n \in \mathbb{N}$, we define $\|A\|_\infty$ by

$$\|A\|_\infty := \sup_{1 \le i \le m} \sup_{1 \le j \le n} |A_i^j|,$$

where A_i^j denotes the (i, j)-component of the matrix A. For a vector $v \in \mathbb{R}^p$, and an index set $T \subset \{1, 2, \ldots, p\}$, we denote the $|T|$-dimensional sub-vector of v restricted by the index set T by v_T, where $|T|$ is the number of elements of the set T. Similarly, for a $p \times p$ matrix A and index sets $T, T' \subset \{1, 2, \ldots, p\}$, we define the $|T| \times |T'|$ sub-matrix $A_{T,T'}$ by

$$A_{T,T'} := (A_i^j)_{i \in T, j \in T'}.$$

3 Preliminaries

Let $\{W_t^1\}_{t \ge 0}, \{W_t^2\}_{t \ge 0}, \ldots$ be independent standard Brownian motions on a probability space $(\Omega, \mathcal{F}, P)$. Define the filtration $\{\mathcal{F}_t\}_{t \ge 0}$ as follows.

$$\mathcal{F}_t := \mathcal{F}_0 \vee \sigma(W_s^j;\ j = 1, 2, \ldots,\ s \in [0, t]), \quad t \ge 0,$$

where $\mathcal{F}_0$ is a σ-field independent of $\{W_t^j\}_{t \ge 0}, j = 1, 2, \ldots$. We consider the following p-dimensional linear stochastic differential equation (1) defined on the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \ge 0}, P)$:

$$X_t = X_0 + \int_0^t \Theta^\top \phi(X_s) ds + \sigma W_t, \quad t \ge 0$$

where $\{W_t\}_{t \ge 0} := \{(W_t^1, \ldots, W_t^p)\}_{t \ge 0}$ is a p-dimensional standard Brownian motion, Θ is a $p \times p$ deterministic matrix, $\sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ is a $p \times p$ diagonal matrix, and $\phi(x) = (\phi_1(x_1), \ldots, \phi_p(x_p))^\top, x = (x_1, \ldots, x_p)^\top$ is a smooth $\mathbb{R}^p$-valued function. Assume that X_0 is $\mathcal{F}_0$-measurable. Note that $\{X_t^i\}_{t \ge 0}$ for each $i = 1, 2, \ldots, p$ satisfies the following equation.

$$X_t^i = X_0^i + \int_0^t \Theta_i^\top \phi(X_s) ds + \sigma_i W_t^i, \quad t \ge 0,$$

where Θ_i is the i-th row of the matrix Θ. In this paper, we consider the estimation problem of the true value (Θ^0, σ^0) of (Θ, σ). Suppose that we can observe the process $\{X_t\}_{t\geq 0}$ at $n+1$ discrete time points:

$$0 =: t_0^n < t_1^n < \ldots < t_n^n, \quad t_k^n = \frac{kt_n^n}{n}, \quad k = 0, 1, \ldots, n.$$

Write T_0^i for the support of the true value Θ_i^0 for every $i \in \{1, 2, \ldots, p\}$, *i.e.*, $T_0^i = \{j : \Theta_{ij}^0 \neq 0\}$. Let S_i be the number of elements in the index set T_0^i. Hereafter, we assume the following high-dimensional and sparse setting for the true matrix Θ^0.

$$p = p_n \gg n, \qquad \sup_{1\leq i<\infty} S_i =: S^* < \infty;$$

note that $S^* > 0$ is a constant which does not depend on n. We use the quasi-likelihood method which is commonly used in this field to estimate the unknown parameters. The quasi-likelihood function is constructed by discretization of the processes by Euler-Maruyama scheme, which is based on the fact that diffusion processes can be locally approximated by Gaussian random variables. See e.g. Yoshida (1992), Genon-Catalot and Jacod (1993) and Kessler (1997) for details. In this model, the quasi-log-likelihood function is given by

$$\sum_{k=1}^{n} \left\{ -\frac{1}{2}\log(2\pi\sigma_i^2\Delta_n) - \frac{|X_{t_k^n}^i - X_{t_{k-1}^n}^i - \Theta_i^T\phi(X_{t_{k-1}^n})\Delta_n|^2}{2\sigma_i^2\Delta_n} \right\},$$

where $\Delta_n := t_k^n - t_{k-1}^n = t_n^n/n$. We write $l_n(\Theta_i, \sigma_i)$ for the normalized quasi-log-likelihood, *i.e.*,

$$l_n(\Theta_i, \sigma_i) := \frac{1}{n\Delta_n}\sum_{k=1}^{n} \left\{ -\frac{1}{2}\log(2\pi\sigma_i^2\Delta_n) - \frac{|X_{t_k^n}^i - X_{t_{k-1}^n}^i - \Theta_i^T\phi(X_{t_{k-1}^n})\Delta_n|^2}{2\sigma_i^2\Delta_n} \right\}.$$

We assume the following conditions.

Assumption 3.1. *(i) It holds that $p_n \to \infty$, that $\log p_n/\sqrt{n\Delta_n} \to 0$, and that $\Delta_n = \Delta n^{-\alpha}$, for some $\alpha \in (1/2, 1)$ and positive constant Δ. Especially, the last condition implies that $n\Delta_n = t_n^n \to \infty$, $\Delta_n \to 0$ and that $n\Delta_n^2 \to 0$ as $n \to \infty$.*

(ii) The functions ϕ_i's are uniformly bounded and satisfy the global Lipschitz condition, i.e., there exist positive constants L and L' such that

$$\sup_{1\leq i<\infty}\sup_{x\in\mathbb{R}} |\phi_i(x)| \leq L$$

and that

$$\sup_{1\leq i<\infty} |\phi_i(x) - \phi_i(y)| \leq L'|x-y|, \quad \forall x, y \in \mathbb{R}.$$

(iii) For every $\nu \geq 1$, there exists a positive constant $\tilde{C}_\nu$ such that

$$\sup_{1\leq i<\infty} \sup_{t\in[0,\infty)} E\left[|X_t^i|^\nu\right] \leq \tilde{C}_\nu.$$

Note that this assumption implies that

$$\sup_{t\in[0,\infty)} E\left[\sup_{1\leq i\leq p_n} |X_t^i|^\nu\right] \leq p_n \tilde{C}_\nu, \quad \forall n \in \mathbb{N}.$$

(iv) There exist some positive constants $K_1,\ K_2,\ K_3$ and K_4 such that

$$K_2 < \inf_{1\leq i<\infty} \inf_{j\in T_0^i} |\Theta_{ij}^0| \leq \sup_{1\leq i,j<\infty} |\Theta_{ij}^0| < K_1,$$

$$K_4 < \inf_{1\leq i<\infty} |\sigma_i^0| \leq \sup_{1\leq i<\infty} |\sigma_i^0| < K_3.$$

(v) For every $i \in \mathbb{N}$, the $\mathbb{R}^{S_i}$-valued process $\{X_{tT_0^i}\}_{t\in[0,T_n]}$ is ergodic for $\Theta = \Theta^0$ and $\sigma = \sigma^0$ with invariant measure μ_0^i.

4 Estimators for diffusion coefficients

It is well-known that we can ignore the influence of drift coefficients when we estimate the diffusion coefficients (see e.g. Yoshida (1992)). We thus define the estimator for σ_i^0 by the solution $\hat{\sigma}_{n,i}$ to the equation

$$\frac{\partial}{\partial \sigma_i} l_n(0, \sigma_i) = 0, \quad i = 1, 2, \ldots, p_n,$$

by letting $\Theta = 0$. Note that $\hat{\sigma}_{n,i}$ can be written explicitly in the following way:

$$\hat{\sigma}_i^2 := \hat{\sigma}_{n,i}^2 = \frac{1}{n\Delta_n} \sum_{k=1}^n |X_{t_k^n}^i - X_{t_{k-1}^n}^i|^2.$$

The next theorem asserts the consistency of $\hat{\sigma}_i$ uniformly in i.

Theorem 4.1. *Under Assumption 3.1, it holds that*

$$\sup_{1\leq i\leq p_n} |\hat{\sigma}_i^2 - (\sigma_i^0)^2| \to^p 0, \quad n \to \infty.$$

Note that Theorem 4.1 and Assumption 3.1 imply that there exists a constant $\tilde{K}_1$ such that

$$\lim_{n\to\infty} P\left(\sup_{1\leq i\leq p_n} \hat{\sigma}_i^{-2} \geq \tilde{K}_1\right) = 0.$$

5 Estimators for drift coefficients

In this section, we define the estimator for Θ_i by plugging $\hat{\sigma}_i$ in quasi-log-likelihood l_n. Hereafter, we write $\psi_n(\Theta_i)$ for the gradient of $l_n(\Theta_i, \hat{\sigma}_i)$ with respect to Θ_i, and V_n^i for Hessian of $-l_n(\Theta_i, \hat{\sigma}_i)$, *i.e.*,

$$\psi_n(\Theta_i) := \frac{1}{n\Delta_n \hat{\sigma}_i^2} \sum_{k=1}^n \phi(X_{t_{k-1}^n})(X_{t_k^n}^i - X_{t_{k-1}^n}^i - \Theta_i^T \phi(X_{t_{k-1}^n})\Delta_n),$$

$$V_n^i := \frac{1}{n\hat{\sigma}_i^2} \sum_{k=1}^n \phi(X_{t_{k-1}^n})\phi(X_{t_{k-1}^n})^\top.$$

Note that the Hessian matrix does not depend on Θ. Define the Dantzig selector type estimator $\hat{\Theta}_{n,i}$ for Θ_i^0 by

$$\hat{\Theta}_{n,i} := \hat{\Theta}_i := \arg\min_{\Theta_i \in \mathcal{C}_n^i} \|\Theta_i\|_1, \quad \mathcal{C}_n^i := \{\Theta_i \in \mathbb{R}^{p_n} : \|\psi_n(\Theta_i)\|_\infty \le \gamma_n^i\},$$

where γ_n^i is a tuning parameter. Hereafter, we assume the following condition about γ_n^i

Assumption 5.1. *γ_n^i satisfies the following equality for some positive constants c^i's which are uniformly bounded in i:*

$$\gamma_n^i = c^i \tilde{\gamma}_n,$$

where $\tilde{\gamma}_n := (\log p_n / n\Delta_n)^{1/4}$.

We define the quantity γ_n by

$$\gamma_n = \sup_{1 \le i \le p_n} \gamma_n^i.$$

Under Assumption 5.1, it is obvious that there exists a constant $c \in (0, \infty)$ such that

$$\frac{\gamma_n}{\tilde{\gamma}_n} = \sup_{1 \le i \le p_n} c^i \le c.$$

Some remarks about the choice of c^i's are described in this talk. In order to prove the l_q consistency of the estimator for every $q \in [1, \infty]$, we need to discuss on the gradient $\psi_n(\Theta_i^0)$ and Hessian matrix V_n^i. Since the main term of $\psi_n(\Theta_i^0)$ is the terminal value of a martingale with finite moment of any order, we can use Bernstein's inequality and maximal inequality for sub-Gaussian variable. Then, we have the following theorem.

Theorem 5.2. *Under Assumptions 3.1 and 5.1, it holds that*

$$\lim_{n \to \infty} P\left(\sup_{1 \le i \le p_n} \|\psi_n(\Theta_i^0)\|_\infty \ge 2\gamma_n \right) = 0.$$

To consider the appropriate condition on the high-dimensional matrix, we introduce the following factors for V_n^i.

Definition 5.3. *For every index set $T \subset \{1,\ 2,\ \cdots,\ p_n\}$ and $h \in \mathbb{R}^{p_n}$, h_T is an $\mathbb{R}^{|T|}$ dimensional sub-vector of h constructed by extracting the components of h corresponding to the indices in T. Define the set C_T by*

$$C_T := \{h \in \mathbb{R}^{p_n} : \|h_{T^c}\|_1 \leq \|h_T\|_1\}.$$

(i) **Compatibility factor**

$$\kappa(T_0^i, V_n^i) := \inf_{0 \neq h \in C_{T_0^i}} \frac{S_i^{\frac{1}{2}} (h^T V_n^i h)^{\frac{1}{2}}}{\|h_{T_0^i}\|_1}.$$

(ii) **Weak cone invertibility factor**

$$F_q(T_0^i, V_n^i) := \inf_{0 \neq h \in C_{T_0^i}} \frac{S_i^{\frac{1}{q}} h^T V_n^i h}{\|h_{T_0^i}\|_1 \|h\|_q}, \quad q \in [1, \infty).$$

$$F_\infty(T_0^i, V_n^i) := \inf_{0 \neq h \in C_{T_0^i}} \frac{(h^T V_n^i h)^{\frac{1}{2}}}{\|h\|_\infty}.$$

(iii) **Restricted eigenvalue**

$$RE(T_0^i, V_n^i) := \inf_{0 \neq h \in C_{T_0^i}} \frac{(h^T V_n^i h)^{\frac{1}{2}}}{\|h\|_2}.$$

We assume that $\kappa(T_0^i, V_n^i)$ satisfies the following condition.

Assumption 5.4. *For every $\epsilon > 0$, there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,*

$$P\left(\inf_{1 \leq i \leq p_n} \kappa(T_0^i, V_n^i) > \delta\right) \geq 1 - \epsilon.$$

Noting that $\|h_{T_0^i}\|_1^q \geq \|h_{T_0^i}\|_q^q$ for all $q \geq 1$, we can see that $\kappa(T_0^i; V_n^i) \leq 2\sqrt{S_i} RE(T_0^i; V_n^i)$, and that $\kappa(T_0^i; V_n^i) \leq F_q(T_0^i; V_n^i)$. So under Assumption 5.4, the two factors $RE(T_0^i; V_n^i)$ and $F_q(T_0^i; V_n^i)$ also satisfy the corresponding conditions. See van de Geer and Bühlmann (2009) for the details of the matrix conditions to deal with the sparsity. The following theorem give the l_q consistency of $\hat{\Theta}_i$ uniformly in i for every $q \in [1, \infty]$.

Theorem 5.5. *Under Assumptions 3.1, 5.1 and 5.4, the following (i)-(iv) hold true.*

(i) *It holds that*

$$\lim_{n \to \infty} P\left(\sup_{1 \leq i \leq p_n} \|\hat{\Theta}_i - \Theta_i^0\|_2^2 \geq \frac{4 \sup_{1 \leq i \leq p_n} \|\Theta_i^0\|_1 \gamma_n}{\inf_{1 \leq i \leq p_n} RE^2(T_0^i, V_n^i)}\right) = 0.$$

In particular, it holds that $\sup_{1 \leq i \leq p_n} \|\hat{\Theta}_i - \Theta_i^0\|_2 \to^p 0$ *as* $n \to \infty$.

(*ii*) *It holds that*

$$\lim_{n\to\infty} P\left(\sup_{1\le i\le p_n} \|\hat{\Theta}_i - \Theta_i^0\|_\infty^2 \ge \frac{4\sup_{1\le i\le p_n}\|\Theta_i^0\|_1\gamma_n}{\inf_{1\le i\le p_n} F_\infty^2(T_0^i, V_n^i)}\right) = 0.$$

In particular, it holds that $\sup_{1\le i\le p_n}\|\hat{\Theta}_i - \Theta_i^0\|_\infty \to^p 0$ *as* $n\to\infty$.

(*iii*) *It holds that*

$$\lim_{n\to\infty} P\left(\sup_{1\le i\le p_n} \|\hat{\Theta}_i - \Theta_i^0\|_1 \ge \frac{8S^*\gamma_n}{\inf_{1\le i\le p_n} \kappa^2(T_0^i, V_n^i)}\right) = 0.$$

In particular, it holds that $\sup_{1\le i\le p_n}\|\hat{\Theta}_i - \Theta_i^0\|_2 \to^p 0$ *as* $n\to\infty$.

(*iv*) *It holds for every* $q\in(1,\infty)$ *that*

$$\lim_{n\to\infty} P\left(\sup_{1\le i\le p_n} \|\hat{\Theta}_i - \Theta_i^0\|_q \ge \frac{4S^{*\frac{1}{q}}\gamma_n}{\inf_{1\le i\le p_n} F_q(T_0^i, V_n^i)}\right) = 0.$$

In particular, it holds that $\sup_{1\le i\le p_n}\|\hat{\Theta}_i - \Theta_i^0\|_q \to^p 0$ *as* $n\to\infty$.

6 Variable selection by the Dantzig selector

6.1 Estimator for the support index set of the drift coefficients

Noting that the rate of convergence l_1 error for the Dantzig selector is γ_n, we can propose the estimator of the support index set T_0^i of the true value Θ_i^0 as follows.

$$\hat{T}_n^i := \{j : |\hat{\Theta}_{ij}| > \gamma_n^i\}.$$

We can prove that $\hat{T}_n^i = T_0^i$ for sufficiently large n with probability tending to 1.

Theorem 6.1. *Under Assumptions 3.1, 5.1 and 5.4, it holds that*

$$\lim_{n\to\infty} P\left(\hat{T}_n^i = T_0^i \text{ for all } i\in\{1,2,\dots,p_n\}\right) = 1.$$

6.2 New estimator for drift coefficients after variable selection

We can construct the new estimator $\hat{\Theta}_i^{(2)}$ by the solution to the next equation

$$\psi_n(\Theta_i)_{\hat{T}_n^i} = 0, \quad \Theta_{i(\hat{T}_n^i)^c} = 0. \tag{2}$$

We will prove the asymptotic normality of the estimator $\hat{\Theta}_{i\hat{T}_n^i}^{(2)}$ for every $i\in\{1,2,\dots,p_n\}$. In order to consider the asymptotic distribution, we assume the following condition about the Fisher information matrix.

Assumption 6.2. *Define the $S_i \times S_i$ matrix $Q^i_{T^i_0,T^i_0}$ by*

$$Q^i_{T^i_0,T^i_0} := \frac{1}{(\sigma^0_i)^2} \int_{\mathbb{R}^{S_i}} \phi(x)_{T^i_0} \phi(x)^\top_{T^i_0} \mu^i_0(dx).$$

It holds that $Q^i_{T^i_0,T^i_0}$ is invertible for every $i = 1, 2, \ldots, p_n$.

Then, we can prove the asymptotic normality of $\hat{\Theta}^{(2)}_{i\hat{T}^i_n}$ in the following sense.

Theorem 6.3. *Under Assumptions 3.1, 5.1, 5.4 and 6.2, it holds for every $i \in \mathbb{N}$ that*

$$\sqrt{n\Delta_n}(\hat{\Theta}^{(2)}_{i\hat{T}^i_n} - \Theta^0_{iT^i_0}) 1_{\{\hat{T}^i_n = T^i_0\}} \to^d N\left(0, \left(Q^i_{T^i_0,T^i_0}\right)^{-1}\right)$$

as $n \to \infty$. Note that for every $i \in \mathbb{N}$, it holds that $i < p_n$ for sufficiently large n.

References

Antoniadis, A., Fryzlewicz, P. and Letué, F. The Dantzig selector in Cox's proportional hazards model. *Scand. J. Stat.* **37**, no.4, 531-552. (2010).

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37**, no. 4, 1705-1732. (2009).

Candés, E. and Tao, T. The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* **35**, no.6, 2313-2351. (2007).

Fan, Y., Gai, Y. and Zhu, L. Asymtotics of Dantzig selector for a general single-index model. *J. Syst. Sci. Complex.* **29**, no.4, 1123-1144. (2016).

Fujimori, K. (2017). Cox's proportional hazards model with a high-dimensional and sparse regression parameter. arXiv:1710.10416[math.ST].

Genon-Catalot, V. and Jacod, J. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**, no.1, 119-151. (1993).

Kessler, M. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, no.2, 211-229. (1997).

Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. Ser. B* **58**, no.1, 267-288. (1996).

van de Geer, S. A. and Bühlmann, P. On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3**, 1360-1392. (2009).

Yoshida, N. Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41**, no.2, 220-242. (1992).