Predictive model selection criteria for Bayesian lasso

Kawano, Shuichi
Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University

Hoshina, Ibuki
九州大学マス・フォア・インダストリ研究所

Matsuda, Kazuki

Konishi, Sadanori

http://hdl.handle.net/2324/1263083
Predictive model selection criteria for Bayesian lasso

Shuichi Kawano, Ibuki Hoshina, Kazuki Matsuda & Sadanori Konishi

MI 2013-12

(Received November 8, 2013)
Predictive model selection criteria for Bayesian lasso

Shuichi Kawano\textsuperscript{1}, Ibuki Hoshina\textsuperscript{2},
Kazuki Matsuda\textsuperscript{2} and Sadanori Konishi\textsuperscript{3}

\textsuperscript{1} Department of Mathematical Sciences, Graduate School of Engineering,
Osaka Prefecture University, 1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan.

\textsuperscript{2} Department of Mathematics, Graduate School of Science and Engineering,
Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.

\textsuperscript{3} Department of Mathematics, Faculty of Science and Engineering, Chuo University,
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.

skawano@ms.osakafu-u.ac.jp  ibu@gug.math.chuo-u.ac.jp
kazuki@gug.math.chuo-u.ac.jp  konishi@math.chuo-u.ac.jp

Abstract: We consider the Bayesian lasso for regression, which is an $L_1$ penalized
regression based on Bayesian approach. In Bayesian theory, a crucial issue is the
specification of prior distributions for parameters, which leads to the selection of
values of hyperparameters included in the prior distributions. In order to select the
values of the hyperparameters, we introduce a model selection criterion by evaluating
the Bayesian predictive distribution for the Bayesian lasso. Several numerical studies
are presented to illustrate the effectiveness of our proposed modeling procedure.

Key Words and Phrases: Bayesian predictive distribution, Information criterion,
Kullback-Leibler information, $L_1$ regularization, Markov chain Monte Carlo.

1 Introduction

In regression analysis, variable selection plays an important role in the extraction of in-
formation from huge-scale datasets with complex structures. However, classical variable
selection procedures such as the best subset selection and forward stepwise selection are
often computationally expensive and extremely unstable because of their inherent dis-
creteness (Breiman, 1996). In addition, traditional estimation methods such as ordinary
least squares and maximum likelihood methods often lead to unstable models when multi-
collinearity or inherent high-dimensionality exists in the data, and hence the resulting
models tend to have poor prediction accuracy.

In order to overcome these problems, various sparse regularization methods, such as
lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), elastic net (Zou and Hastie, 2005),
adaptive lasso (Zou, 2006), and minimax concave penalty (Zhang, 2010), have been pro-
posed. These procedures enable us to perform simultaneous variable selection and stable
parameter estimation by selecting appropriate values of the tuning parameters that de-
termine the amount of penalties. It is well known that penalty functions in regularization
correspond to the specification of prior distributions in Bayesian models; that is, the reg-
ularization methods can be regarded as a Bayesian approach. For example, a coefficient
estimator for lasso could be interpreted as a posterior mode under independent Laplace
prior distributions. From this perspective, Park and Casella (2008) and Hans (2009) pro-
posed the Bayesian lasso, which is a Bayesian treatment for lasso. Subsequently, sevaryl
variations and extensions of the Bayesian lasso have been proposed and these methods
have been used in various fields (e.g., Kabán, 2007; Kyung et al., 2010; Li and Lin, 2010;
Mutshinda and Sillanpää, 2010; Li et al., 2011; Leng et al., 2013).

A crucial issue for the Bayesian lasso is the choice of values of hyperparameters in-
cluded in prior distributions, which corresponds to the selection of values of tuning pa-
rameters included in the regularization term. Park and Casella (2008) proposed that the
values of hyperparameters are determined by the use of a hierarchical or empirical Bayes
approach. Hans (2010) proposed a variable selection procedure that can treat model un-
certainty based on the marginal likelihood. The deviance information criterion (DIC)
proposed by Spiegelhalter et al. (2002), which is one of the most popular model selection
criteria from the viewpoints of Bayesian approaches, and other Bayesian variable selec-
tion methods (e.g., O’Hara and Sillanpää, 2009) would also be applicable for solving the
problem. However, there have been no studies on the choice of values of hyperparameters
by evaluating the Bayesian predictive distribution for the Bayesian lasso, since obtaining
this distribution is difficult, as the prior distributions are not conjugated for the likelihood
In this paper, we present a model selection criterion by evaluating the Bayesian predictive distribution for the Bayesian lasso from the viewpoints of an information-theoretic approach. First, we obtain an approximated prior distribution by approximating the Laplace prior distribution by the normal prior distribution based on the Kullback-Leibler information. Using this approximated prior distribution, we can analytically obtain the formula of the Bayesian predictive distribution for the Bayesian lasso. In order to derive our proposed criterion, we employ the idea of the predictive information criterion (PIC) proposed by Kitagawa (1997). Several numerical studies are conducted to investigate the performances of our proposed procedure.

The rest of this paper is organized as follows. Section 2 describes the Bayesian lasso. In Section 3, we introduce a model selection criterion for evaluating the Bayesian predictive distribution for the Bayesian lasso. Monte Carlo simulations and real data analysis are conducted to examine the performances of our proposed procedure and to compare it with existing methods in Section 4. Concluding remarks are given in Section 5.

2 Bayesian lasso

2.1 Preliminaries

We consider the linear regression model

\[ y = \beta_0 1_n + X \beta + \varepsilon, \]

where \( y = (y_1, \ldots, y_n)^T \) is an \( n \)-dimensional response vector, \( 1_n \) is an \( n \)-dimensional vector whose elements are all one, \( X = (x_1, \ldots, x_n)^T \) is an \( n \times p \) design matrix, \( \beta_0 \) is an intercept, \( \beta = (\beta_1, \ldots, \beta_p)^T \) is a \( p \)-dimensional coefficient vector and \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \) is an \( n \)-dimensional error vector distributed as \( N_n(0, \sigma^2 I_n) \). Here, \( x_i = (x_{i1}, \ldots, x_{ip})^T \) \((i = 1, \ldots, n)\) denotes a \( p \)-dimensional covariate vector, \( \sigma (> 0) \) is an unknown parameter and \( I_n \) is an \( n \times n \) identity matrix. Without loss of generality, we assume that the response
vector is centered and that the design matrix $X$ is standardized:

$$
\sum_{i=1}^{n} y_i = 0, \quad \sum_{i=1}^{n} x_{ij} = 0, \quad \sum_{i=1}^{n} x_{ij}^2 = n, \quad j = 1, \ldots, p.
$$

From this assumption, Equation (1) is replaced by

$$
y = X\beta + \varepsilon.
$$

Since the error vector $\varepsilon$ is distributed as a multivariate normal distribution $N_n(0, \sigma^2 I_n)$, we have a probability density function for the response $y_i$ ($i = 1, \ldots, n$) in the form

$$
f(y_i|x_i; \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(y_i - x_i^T\beta)^2}{2\sigma^2} \right], \quad i = 1, \ldots, n.
$$

This leads to the log-likelihood function

$$
\sum_{i=1}^{n} \log f(y_i|x_i; \beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i^T\beta)^2.
$$

Hereinafter, we represent the probability density function $f(y_i|x_i; \beta, \sigma^2)$ as $f(y_i|\beta, \sigma^2)$, for simplicity.

### 2.2 Lasso

By maximizing the log-likelihood function in Equation (5) with respect to the parameters $\beta$ and $\sigma^2$, we obtain the maximum likelihood estimator for the parameters. However, the maximum likelihood method (MLE) does not enable us to perform parameter estimation and variable selection simultaneously; thus MLE cannot produce sparse solutions for coefficient parameters. We therefore estimate the parameters $\beta$ and $\sigma^2$ by maximizing the log-likelihood function with $L_1$ penalty:

$$
\max_{\beta, \sigma^2} \left\{ \sum_{i=1}^{n} \log f(y_i|\beta, \sigma^2) - n\lambda \sum_{j=1}^{p} |\beta_j| \right\},
$$

where $\lambda (> 0)$ is a regularization parameter. This estimation procedure is called lasso (Tibshirani, 1996).

Equation (6) is a concave optimization problem, and hence there exists in unique solution for these parameters. Meanwhile, the solution of lasso is not usually expressed in a
closed form since the second term in Equation (6) is non-differentiable. Therefore, various estimation algorithms for lasso have been developed to derive the estimator numerically, e.g., LARS (least angle regression) algorithm by Efron et al. (2004) and coordinate decent algorithm by Friedman et al. (2008).

2.3 Bayesian lasso

In estimating the parameters $\beta$ and $\sigma^2$, we take a Bayesian approach by specifying prior distributions for the coefficient parameter $\beta$ and variance $\sigma^2$. This procedure was proposed by Park and Casella (2008) and is called the Bayesian lasso.

In order to treat lasso in terms of Bayesian theory, we assume a conditional Laplace prior on $\beta$ of the form

$$
\pi(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{n\lambda}{2\sqrt{\sigma^2}} \exp \left[ -\frac{n\lambda|\beta_j|}{\sqrt{\sigma^2}} \right] \tag{7}
$$

and the noninformative scale-invariant marginal prior $\pi(\sigma^2) = 1/\sigma^2$ or inverse-gamma prior $\pi(\sigma^2) = IG(\nu_0/2, \eta_0/2)$ on $\sigma^2$, where $\nu_0/2$ is a shape parameter, $\eta_0/2$ is a scale parameter, and both these parameters are positive. As an alternative specification for Equation (7), Park and Casella (2008) proposed the following hierarchical representation

$$
\pi(\beta|\sigma^2, \tau_1^2, \ldots, \tau_p^2) = \prod_{j=1}^{p} \frac{n}{2\sigma^2} \exp \left[ -\frac{n^2\beta_j^2}{2\sigma^2\tau_j^2} \right], \tag{8}
$$

$$
\pi(\tau_1^2, \ldots, \tau_p^2) = \prod_{j=1}^{p} \frac{\lambda^2}{2} \exp \left[ -\frac{\lambda^2\tau_j^2}{2} \right]. \tag{9}
$$

The specification of the prior distributions is based on representing the Laplace distribution as a scale mixture of normals. For more details, we refer to Andrews and Mallows (1974) and Park and Casella (2008).

The formulation enables us to implement the Gibbs sampler for $\beta, \sigma^2$ and $\tau_1^2, \ldots, \tau_p^2$. Assuming an inverse-gamma prior distribution $\pi(\sigma^2) = IG(\nu_0/2, \eta_0/2)$ on $\sigma^2$, the full conditional posterior distributions of $\beta$, $\sigma^2$ and $1/\tau_j^2$ ($j = 1, \ldots, p$) are, respectively, given
by

$$\beta \mid y, X, \sigma^2, \tau_1^2, \ldots, \tau_p^2 \sim N(A^{-1}X^T y, \sigma^2 A^{-1}),$$

$$A = X^T X + n^2 D^{-1}_\tau, \ D_\tau = \text{diag}(\tau_1^2, \ldots, \tau_p^2),$$

$$\sigma^2 \mid y, X, \beta, \tau_1^2, \ldots, \tau_p^2 \sim IG\left(\frac{\nu_1}{2}, \frac{\eta_1}{2}\right),$$

$$\nu_1 = n + p + \nu_0, \ \eta_1 = ||y - X\beta||^2 + n^2 \beta^T D^{-1}_\tau \beta + \eta_0,$$

$$\frac{1}{\tau_j} \beta_j, \sigma^2, \lambda \sim IGauss(\mu', \lambda'),$$

$$\mu' = \sqrt{\frac{\lambda^2 \sigma^2}{n^2 \beta_j^2}}, \ \lambda' = \lambda^2, \ j = 1, \ldots, p,$$

where $IGauss(\mu, \lambda)$ denotes the inverse Gaussian distribution with density function

$$f(x \mid \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^{-3/2}}} \exp\left[-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right], \ x > 0.$$
Table 1: Sparse algorithm (Hoshina, 2012).

Sparse algorithm

1. Estimate the coefficient vector $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$

2. $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_p)^T \leftarrow \hat{\beta}$

3. For $j = 1, \ldots, p$,
   
   set $\tilde{\beta}_j \leftarrow 0$

3.1 if $g(\hat{\beta}, \hat{\xi}, y) \geq g(\tilde{\beta}, \hat{\xi}, y)$ then $\hat{\beta}_j \leftarrow \tilde{\beta}_j$

3.2 else $\hat{\beta}_j \leftarrow \tilde{\beta}_j$

where $g(\beta, \xi, y) = \log f(y|\beta, \xi) + \log \pi(\beta, \xi)$,

$f(y|\beta, \xi)$ is a likelihood, $\pi(\beta, \xi)$ is a prior on $(\beta, \xi)$,

and $\hat{\xi}$ is point estimates of the parameter vector

$\xi = (\sigma^2, \tau_1^2, \ldots, \tau_p^2)^T$.

Remarks Hyperparamters to be determined include $\lambda$ in the prior distribution on $\beta$
and $\nu_0, \eta_0$ in the prior distribution on $\sigma^2$. In this paper, we focus on the selection of the
value of only hyperparamter $\lambda$, since it is difficult to optimize values of all hyperparamters
$(\lambda, \nu_0, \eta_0)$ simultaneously in terms of computational times. We leave this problem as a
future research topic.

3 Model selection criterion

Kitagawa (1997) proposed the predictive information criterion (PIC) by evaluating the
Bayesian predictive distribution. The PIC is, in general, given by

$$\text{PIC} = -2 \log h(y|y) + 2B_p, \quad (10)$$

where $h(z|y)$ is a Bayesian predictive distribution of the form

$$h(z|y) = \int f(z|\beta, \sigma^2)\pi(\beta, \sigma^2|y)d\beta d\sigma^2, \quad (11)$$

in which $z = (z_1, \ldots, z_n)^T$ is an $n$-dimensional future observation, $f(z|\beta, \sigma^2) = \prod_{i=1}^n f(z_i|\beta, \sigma^2)$
and \( \pi(\beta, \sigma^2 | y) \) is the joint posterior distribution

\[
\pi(\beta, \sigma^2 | y) = \frac{f(y | \beta, \sigma^2)\pi(\beta, \sigma^2)}{\int f(y | \beta, \sigma^2)\pi(\beta, \sigma^2)d\beta d\sigma^2}
\]

\[
= \frac{f(y | \beta, \sigma^2)\pi(\beta^2|\sigma^2)\pi(\sigma^2)}{\int f(y | \beta, \sigma^2)\pi(\beta|\sigma^2)\pi(\sigma^2)d\beta d\sigma^2}, \tag{12}
\]

and where \( B_p \) is the bias term given by

\[
B_p = E_{q(y)} \left[ \log h(y | y) - E_{q(z)} \left[ \log h(z | y) \right] \right]. \tag{13}
\]

with \( q(\cdot) \) be the true distribution that generates the data.

We now consider PIC for the Bayesian lasso. In order to derive PIC, we first need to obtain the Bayesian predictive distribution in Equation (11). In the Bayesian lasso, the prior distribution is formulated by Equation (7). However, it is difficult to obtain the predictive distribution \( h(z | y) \) based on these priors analytically, since we cannot analytically describe the form of the posterior distribution. This problem arises from the fact that the prior distribution \( \pi(\beta | \sigma^2) \) is not a conjugated prior for the likelihood function. In Section 3.1, we approximate the prior distribution \( \pi(\beta | \sigma^2) \) by a conjugated prior distribution (a normal prior distribution) for the likelihood function.

### 3.1 Approximated prior distribution

Let \( f(\beta) \) be the Laplace distribution given by

\[
f(\beta) = \frac{n\lambda}{2\sqrt{\sigma^2}} \exp \left( -\frac{n\lambda|\beta|}{\sqrt{\sigma^2}} \right), \tag{14}
\]

and \( g(\beta | \alpha^2) \) be the normal distribution given by

\[
g(\beta | \alpha^2) = \frac{1}{\sqrt{2\pi\alpha^2}} \exp \left( -\frac{\beta^2}{2\alpha^2} \right), \tag{15}
\]

where \( \alpha \) is positive.

Our aim is to find the normal distribution that is the closest to the Laplace distribution. Here, we measure the closeness between the distributions in terms of the Kullback-Leibler information (Kullback and Leibler, 1951). That is, we determine the normal distribution \( g(\beta | \hat{\alpha}^2) \), where \( \hat{\alpha}^2 \) is an estimator of \( \alpha^2 \), such that the Kullback-Leibler information
between the distributions \( f(\beta) \) and \( g(\beta|\alpha^2) \):

\[
\text{KL}(f, g) = \int_{-\infty}^{\infty} f(\beta) \log \frac{f(\beta)}{g(\beta|\alpha^2)} d\beta
\]

(16)

is minimized with respect to the parameter \( \alpha^2 \). Therefore, we can obtain the following theorem.

**Theorem 1.** The minimum of the Kullback-Leibler information in Equation (16) attains at \( \hat{\alpha}^2 = 2(\sqrt{\sigma^2}/n\lambda)^2 \).

**Proof.** The Kullback-Leibler information between \( f(\beta) \) and \( g(\beta|\alpha) \) is calculated as

\[
\text{KL}(f, g) = \log(n\lambda) - \log(2\sqrt{\sigma^2}) + \frac{1}{2} \log(2\pi\alpha^2) - 1 + \frac{1}{\alpha^2} \left( \frac{\sqrt{\sigma^2}}{n\lambda} \right)^2.
\]

(17)

A minimizer of Equation (17) is obtained from \( \frac{\partial \text{KL}(f, g)}{\partial \alpha^2} = 0 \). We then derive \( \hat{\alpha}^2 = 2(\sqrt{\sigma^2}/n\lambda)^2 \).

From this result, we approximate the Laplace distribution \( f(\beta) \) by the normal distribution \( g(\beta|\hat{\alpha}^2) \); we have

\[
\pi(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{n\lambda}{2\sqrt{\sigma^2}} \exp \left[ -\frac{n\lambda|\beta_j|}{2\sqrt{\sigma^2}} \right] \approx \tilde{\pi}(\beta|\sigma^2) = \prod_{j=1}^{p} \frac{n\lambda}{\sqrt{2\pi}(2\sigma^2)} \exp \left[ -\frac{n^2\lambda^2\beta_j^2}{2(2\sigma^2)} \right].
\]

(18)

Note that the approximated distribution \( \tilde{\pi}(\beta|\sigma^2) \) can be regarded as the closest to the Laplace distribution \( \pi(\beta|\sigma^2) \) in terms of the Kullback-Leibler information.

**Remarks** As the measure of the closeness between the Laplace distribution in Equation (14) and the normal distribution in Equation (15), we considered the Kullback-Leibler information. We can, however, employ various measures of the closeness instead of the Kullback-Leibler information.

For example, the \( L^2 \) distance

\[
L^2(f, g) = \int_{-\infty}^{\infty} \{ f(\beta) - g(\beta|\alpha^2) \}^2 d\beta
\]

(19)

is available as the measure of the closeness. The quantity is calculated as

\[
L^2(f, g) = \frac{n\lambda}{4\sqrt{\sigma^2}} - \frac{2n\lambda}{\sqrt{\sigma^2}} \exp \left( \frac{\alpha^2 n^2 \lambda^2}{2\sigma^2} \right) \Phi \left( -\sqrt{\frac{\alpha^2 n^2 \lambda^2}{\sigma^2}} \right) + \frac{1}{2\sqrt{\pi\alpha^2}},
\]

(20)
where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Then, the estimator $\hat{\alpha}^2$ is obtained by solving $\partial L^2(f, g)/\partial \alpha^2 = 0$, that is,

$$
\frac{1}{4\sqrt{\pi(\alpha^2)^{3/2}}} + \frac{n\lambda}{\sqrt{\sigma^2}} \left[ \frac{n^2 \alpha^2}{\sigma^2} \Phi \left( -\sqrt{\frac{\alpha^2}{\sigma^2} n\lambda} \right) - \frac{n\lambda}{\sqrt{\alpha^2\sigma^2}} \phi \left( -\sqrt{\frac{\alpha^2}{\sigma^2} n\lambda} \right) \right] \exp \left( \frac{\alpha^2 n^2 \lambda^2}{2\sigma^2} \right) = 0
$$

with respect to the parameter $\alpha^2$. Here, $\phi(\cdot)$ is the probability distribution function of the standard normal distribution. Since the estimator $\hat{\alpha}^2$ cannot be obtained analytically, we use some numerical algorithms in order to estimate the parameter.

It is interesting to compare performances of the approximated distribution based on the Kullback-Leibler information with those based on the $L^2$ distance. Since our aim is, however, to provide the approximated Laplace distribution based on the Kullback-Leibler information, we will discuss this problem in another paper.

### 3.2 Bayesian predictive distribution for Bayesian lasso

Assuming the approximated prior distribution $\tilde{\pi}(\beta|\sigma^2)$ on $\beta$ and an inverse gamma distribution $\pi(\sigma^2) = IG(\nu_0/2, \eta_0/2)$ on $\sigma^2$ as in Section 2.3, we derive the joint prior distribution $\pi(\beta, \sigma^2)$ of the form

$$
\pi(\beta, \sigma^2) = \pi(\beta|\sigma^2)\pi(\sigma^2) \approx \tilde{\pi}(\beta|\sigma^2)\pi(\sigma^2).
$$

From the prior distribution and Bayes’ rule, the joint posterior distribution can be expressed as

$$
\pi(\beta, \sigma^2|y) = \pi_1(\beta|\sigma^2, y)\pi_2(\sigma^2|y), \tag{22}
$$

where each posterior distribution is given by

$$
\pi_1(\beta|\sigma^2, y) = N(\hat{\beta}_n, \sigma^2 A_n), \quad \pi_2(\sigma^2|y) = IG \left( \frac{\nu_n}{2}, \frac{\hat{\eta}_n}{2} \right). \tag{23}
$$

Here,

$$
A_n = \left( X^T X + \frac{n^2 \lambda^2}{2} I_p \right)^{-1}, \quad \hat{\beta}_n = A_n X^T y, \quad \nu_n = n + \nu_0, \quad \hat{\eta}_n = \eta_0 + y^T y - \hat{\beta}_n^T A_n^{-1} \hat{\beta}_n.
$$

Note that if the prior distribution $\pi(\beta|\sigma^2)$ in Equation (7) is used instead of the approximated prior distribution $\tilde{\pi}(\beta|\sigma^2)$, we cannot obtain the posterior distribution $\pi_1(\beta|\sigma^2, y)$. 

10
Using the posterior distributions, we obtain the Bayesian predictive distribution for the Bayesian lasso given by
\[
h(z|y) = \int f(z|\beta, \sigma^2) \pi(\beta, \sigma^2|y) d\beta d\sigma^2
\]
\[
= \frac{\Gamma \left( \frac{n + \nu_n}{2} \right)}{\Gamma \left( \frac{\nu_n}{2} \right)} \frac{|\Sigma|^{-1/2}}{(\pi \nu_n)^{n/2}} \left[ 1 + \frac{1}{\nu_n} (z - X\hat{\beta}_n)^T \hat{\Sigma}^{-1} (z - X\hat{\beta}_n) \right]^{-(n + \nu_n)/2},
\]
where  \( \hat{\Sigma} = (\hat{\eta}_n/\nu_n)(X^T_n X + I_n) \) and \( \Gamma(\cdot) \) is the Gamma function. This predictive distribution is an \( n \)-dimensional \( t \)-distribution with \( \nu_n \) degrees of freedom.

### 3.3 Proposed criterion

Next, we need to calculate the bias term in Equation (13), since the Bayesian predictive distribution in Equation (24) is obtained in Section 3.2. It is, however, difficult to calculate the bias term analytically, because the Bayesian predictive distribution \( h(z|y) \) in Equation (24) is an \( n \)-dimensional \( t \)-distribution. Hence, we approximate the distribution \( h(z|y) \) by a normal distribution \( f(z|\tilde{\beta}, \tilde{\sigma}^2) \) in the form
\[
h(z|y) = f(z|\tilde{\beta}, \tilde{\sigma}^2) \left\{ 1 + O_p(n^{-1}) \right\},
\]
where \( \tilde{\beta} \) and \( \tilde{\sigma}^2 \) are, respectively, given by
\[
\tilde{\beta} = \left( X^T X + \frac{n^2 \lambda^2}{2} I_p \right)^{-1} X^T y,
\]
\[
\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})^T (y - X\tilde{\beta}) + \frac{n^2 \lambda^2}{2} \tilde{\beta}^T \tilde{\beta} + \eta_0}{n + p + \nu_n + 2}.
\]
This approximation is based on the Laplace approximation (Tierney and Kanade, 1986).

For details of this approximation, we refer to Konishi and Kitagawa (2008).

For the approximated predictive distribution \( f(z|\tilde{\beta}, \tilde{\sigma}^2) \), we define an approximated predictive information criterion (aPIC) as follows:
\[
aPIC = -2 \log h(y|y) + 2B_p^*,
\]
where the approximated bias term \( B_p^* \) is given by
\[
B_p^* = E_q(y) \left[ \log f(y|\tilde{\beta}, \tilde{\sigma}^2) - E_q(z) [\log f(z|\tilde{\beta}, \tilde{\sigma}^2)] \right]
\]
\[
\approx -\frac{1}{2\tilde{\sigma}^2} \left[ E_q(y) [(y - X\tilde{\beta})^T (y - X\tilde{\beta}) - E_q(z) [(z - X\tilde{\beta})^T (z - X\tilde{\beta})]] \right].
\]

11
Using the result of Kim et al. (2012), we can calculate the approximated bias term as

\[ B_p^* \approx \left( \frac{\sigma^{*2}}{\hat{\sigma}^2} \right) \text{tr} \left[ X \left( X^T X + \frac{n^2 \lambda^2}{2} I_p \right)^{-1} X^T \right], \tag{28} \]

where \( \sigma^{*2} \) is a specific value such that \( q(z) = f(z|\beta^*, \sigma^{*2}) \). For more details of the derivations, we refer to Kim et al. (2012).

Then we derive aPIC in the form

\[
\text{aPIC} = -2 \log \Gamma \left( \frac{n + \nu_n}{2} \right) + 2 \log \Gamma \left( \frac{\nu_n}{2} \right) + n \log(\pi \nu_n) + \log \left| \hat{\Sigma} \right| \\
+ (n + \nu_n) \log \left[ 1 + \frac{1}{\nu_n} (y - X \hat{\beta}_n)^T \hat{\Sigma}^{-1} (y - X \hat{\beta}_n) \right] \\
+ 2 \left( \frac{\sigma^{*2}}{\hat{\sigma}^2} \right) \text{tr} \left[ X \left( X^T X + \frac{n^2 \lambda^2}{2} I_p \right)^{-1} X^T \right]. \tag{29} \]

Since the value of \( \sigma^{*2} \) is generally unknown, we replace \( \sigma^{*2} \) by the mode of the posterior distribution \( \hat{\sigma}^2 \), and have

\[
\text{aPIC} = -2 \log \Gamma \left( \frac{n + \nu_n}{2} \right) + 2 \log \Gamma \left( \frac{\nu_n}{2} \right) + n \log(\pi \nu_n) + \log \left| \hat{\Sigma} \right| \\
+ (n + \nu_n) \log \left[ 1 + \frac{1}{\nu_n} (y - X \hat{\beta}_n)^T \hat{\Sigma}^{-1} (y - X \hat{\beta}_n) \right] \\
+ 2 \text{tr} \left[ X \left( X^T X + \frac{n^2 \lambda^2}{2} I_p \right)^{-1} X^T \right]. \tag{30} \]

The value of the hyperparameter \( \lambda \) is selected as the minimizer of aPIC in Equation (30).

**Remarks** Although above \( \sigma^{*2} \) is replaced by the posterior mode \( \hat{\sigma}^2 \), this is not a general recommendation. For example, it is possible to use the posterior mean or another statistics as the value of \( \sigma^{*2} \), and hence it is interesting to investigate which values are most useful. We consider this problem as a future research topic.

### 3.4 Other selection methods

#### 3.4.1 Deviance information criterion

Spiegelhalter et al. (2002) introduced a measure for the effective number of parameters in a model from a Bayesian viewpoint, using an information-theoretic argument. The measure is defined by

\[
p_D = -2 E_{z(\beta, \sigma^2|y)} \left[ \log f(y|\beta, \sigma^2) \right] + 2 \log f(y|\bar{\beta}, \bar{\sigma}^2), \tag{31} \]
where $\bar{\beta}$ and $\bar{\sigma}^2$ are the posterior means defined by $\bar{\beta} = E_{\pi(\beta, \sigma^2)}[\beta]$ and $\bar{\sigma}^2 = E_{\pi(\sigma^2)}[\sigma^2]$, respectively.

Based on this measure, Spiegelhalter et al. (2002) proposed a deviance information criterion (DIC) in the form

$$DIC = -2 \log f(y|\bar{\beta}, \bar{\sigma}^2) + 2p_D. \quad (32)$$

The optimal value of hyperparameter $\lambda$ is chosen by selecting the one that minimizes the value of DIC. Note that DIC is widely used in various fields of research including statistical science, ecology, and marketing (e.g., Brady et al., 2004; Pieters and Wedel, 2004; Martin et al., 2005; Bolker et al., 2009).

### 3.4.2 Full Bayesian approach

An alternative method for choosing the hyperparameter $\lambda$ is a fully Bayesian approach. Park and Casella (2008) assumed the class of gamma prior distributions on $\sigma^2$ given by

$$\text{Gamma}(\sigma^2|\delta, \lambda^2) = \frac{\delta^\lambda}{\Gamma(\lambda)} (\sigma^2)^{\lambda-1} \exp [-\delta \lambda^2], \quad (33)$$

where $\delta$ and $\lambda$ are adjusted parameters with positive values. In our numerical examples, we set $\delta = 0.001$.

The specification in Equation (33) has some attractive properties. For example, the prior distribution on $\lambda^2$ in Equation (33) enables us to easily implement the Gibbs sampler. For more details of other properties, we refer to Park and Casella (2008).

### 4 Numerical studies

#### 4.1 Monte Carlo simulations

The performances of our proposed method were investigated through simulation studies. We generated data according to the following linear regression model

$$y = \mathbf{x}^T \beta^* + \varepsilon, \quad (34)$$

where $\beta^*$ is a $p$-dimensional true coefficient vector, $\varepsilon \sim N(0, \sigma^2)$, and $\mathbf{x}$ is generated from a multivariate normal distribution with mean vector $\mathbf{0}_p$ and covariance matrix $\Sigma$. The
detailed structure of the covariance matrix here is given below. In this simulation, we considered four settings inspired by Tibshirani (1996) as follows:

- **Example 1:** In this example we simulated 200 data sets with 20, 50, or 100 observations. Here, we set \( \beta^* = (3, 1.5, 0, 0, 2, 0, 0, 0)^T \), \( \sigma = 3 \), and the correlation between \( x_i \) and \( x_j \) was \( 0.5^{|i-j|} \).

- **Example 2:** The second example is the same as Example 1, but with \( \beta^* = 0.85 \cdot 1_8 \).

- **Example 3:** The model is the same as Example 1, but with \( \beta^* = (5, 0_7^T)^T \) and \( \sigma = 2 \).

- **Example 4:** In this example we simulated 200 data sets with 100, 200, or 500 observations. Here, we set \( \beta^* = (0_{10}^T, 2_{10}^T, 0_{10}^T, 0_{10}^T)^T \), \( \sigma = 15 \), and the correlation between \( x_i \) and \( x_j \) was \( 0.5 \) (\( i \neq j \)).

In all examples, 2,000 samples from the MCMC simulation were used for estimating the parameters, where the first 1,000 samples were discarded as burn-in. In addition, we confirmed the convergence of the Markov chain simulations by using R.hat (Gelman and Rubin, 1992); the values were close to one. The hyperparameter \( \lambda \) was tested for 200 values; \( \lambda_i = \lambda_{\text{min}} \cdot \exp[(\log \lambda_{\text{max}} - \log \lambda_{\text{min}}) \cdot (i/200)] \) \( (i = 1, \ldots, 200) \), where \( \lambda_{\text{max}} \) is such that all coefficient parameters are zero and \( \lambda_{\text{min}} \) is \( 10^{-4} \) when \( n = 20 \) and \( 10^{-4}/n \) when \( n \) is larger than 50.

The performances of our proposed procedure were evaluated in terms of three accuracies; variable selection, estimation, and prediction accuracies. As the variable selection accuracy, we employed the true positive rate (TPR), true negative rate (TNR), and true sign rate (TSR), respectively, defined by

\[
TPR = \frac{1}{200} \sum_{k=1}^{200} \frac{\left| \left\{ j : \hat{\beta}_j^{(k)} \neq 0 \land \beta_j^* \neq 0 \right\} \right|}{\left| \left\{ j : \beta_j^* \neq 0 \right\} \right|},
\]

\[
TNR = \frac{1}{200} \sum_{k=1}^{200} \frac{\left| \left\{ j : \hat{\beta}_j^{(k)} = 0 \land \beta_j^* = 0 \right\} \right|}{\left| \left\{ j : \beta_j^* = 0 \right\} \right|}.
\]
TSR = \frac{1}{200} \sum_{k=1}^{200} \left\{ j : \text{sign}(\hat{\beta}_j^{(k)}) = \text{sign}(\beta_j^*) \right\},

where \( \hat{\beta}^{(k)} = (\hat{\beta}_1^{(k)}, \ldots, \hat{\beta}_p^{(k)})^T \) is the estimated coefficient vector for the \( k \)-th data set, and \( |\{\ast\}| \) is the number of elements included in a set \{\ast\}. The estimation and prediction accuracies are determined by MSE and PSE as follows;

\[
\text{MSE} = \frac{1}{200} \sum_{k=1}^{200} (\hat{\beta}^{(k)} - \beta^*)^T \Sigma(\hat{\beta}^{(k)} - \beta^*),
\]

\[
\text{PSE} = \frac{1}{200} \sum_{k=1}^{200} \left\{ \frac{1}{n} \|\hat{y}^{(k)} - \tilde{y}^{(k)}\|^2 \right\},
\]

where \( \hat{y}^{(k)} = x^{(k)T} \hat{\beta}^{(k)} \) is the predictor for the \( k \)-th data set, and \( \hat{y}^{(k)} \) is a future observation generated from the model in Equation (34).

For each example, we compared seven procedures; aPIC (proposed procedure), aPIC + SA (aPIC with the sparse algorithm proposed by Hoshina (2012)), DIC, DIC + SA, Blasso (fully Bayesian procedure for the Bayesian lasso proposed by Park and Casella (2008)), Blasso + SA, and Lasso. Here, except for Lasso, the values of the hyperparameters \( \nu_0 \) and \( \eta_0 \) involved in the prior distribution on \( \sigma^2 \) were set to 0.001. The tuning parameter in Lasso was selected by 10-fold cross-validation.

Tables 2 and 3 summarize the simulation results. We observe that aPIC has better estimation and prediction accuracies than other methods. While DIC+SA and Lasso have better performances in terms of TPR, TNR is not better for these two procedures. In Example 4, TPR and TSR in aPIC+SA is superior to those in competitors. From these results, we conclude that our proposed procedure, aPIC and aPIC+SA, may be equally or more useful than competitors.

In most cases, our proposed procedure has better performances in terms of MSE and PSE with increased amount of sample sizes \( n \). However, even if sample sizes \( n \) are large, the results do not necessarily affect the performances of model selection in terms of TPR, TNR and TSR.
Table 2: The results of Example 1 and Example 2.

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 20$</td>
<td>$n = 20$</td>
</tr>
<tr>
<td></td>
<td>TPR</td>
<td>TNR</td>
</tr>
<tr>
<td>aPIC</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>aPIC+SA</td>
<td>0.81</td>
<td>0.62</td>
</tr>
<tr>
<td>DIC</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DIC+SA</td>
<td>0.90</td>
<td>0.43</td>
</tr>
<tr>
<td>Blasso</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Blasso+SA</td>
<td>0.65</td>
<td>0.73</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.90</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>$n = 50$</td>
<td>$n = 50$</td>
</tr>
<tr>
<td></td>
<td>TPR</td>
<td>TNR</td>
</tr>
<tr>
<td>aPIC</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>aPIC+SA</td>
<td>0.99</td>
<td>0.58</td>
</tr>
<tr>
<td>DIC</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DIC+SA</td>
<td>0.99</td>
<td>0.42</td>
</tr>
<tr>
<td>Blasso</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Blasso+SA</td>
<td>0.98</td>
<td>0.51</td>
</tr>
<tr>
<td>Lasso</td>
<td>1.00</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td></td>
<td>TPR</td>
<td>TNR</td>
</tr>
<tr>
<td>aPIC</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>aPIC+SA</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>DIC</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DIC+SA</td>
<td>1.00</td>
<td>0.47</td>
</tr>
<tr>
<td>Blasso</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Blasso+SA</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>Lasso</td>
<td>1.00</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Table 3: The results of Example 3 and Example 4.

<table>
<thead>
<tr>
<th>Example 3</th>
<th>Example 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n = 20</strong></td>
<td><strong>n = 100</strong></td>
</tr>
<tr>
<td>TPR</td>
<td>TNR</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>aPIC</td>
<td>1.00</td>
</tr>
<tr>
<td>aPIC+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>DIC</td>
<td>1.00</td>
</tr>
<tr>
<td>DIC+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>Blasso</td>
<td>1.00</td>
</tr>
<tr>
<td>Blasso+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>Lasso</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>n = 50</strong></th>
<th><strong>n = 200</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>TNR</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>aPIC</td>
<td>1.00</td>
</tr>
<tr>
<td>aPIC+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>DIC</td>
<td>1.00</td>
</tr>
<tr>
<td>DIC+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>Blasso</td>
<td>1.00</td>
</tr>
<tr>
<td>Blasso+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>Lasso</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>n = 100</strong></th>
<th><strong>n = 500</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>TNR</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>aPIC</td>
<td>1.00</td>
</tr>
<tr>
<td>aPIC+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>DIC</td>
<td>1.00</td>
</tr>
<tr>
<td>DIC+SA</td>
<td>1.00</td>
</tr>
<tr>
<td>Blasso</td>
<td>1.00</td>
</tr>
<tr>
<td>Blasso+SA</td>
<td>0.99</td>
</tr>
<tr>
<td>Lasso</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 4: The numbers of samples and predictors for real datasets.

<table>
<thead>
<tr>
<th></th>
<th>diabetes</th>
<th>Boston housing</th>
<th>Parkinson</th>
<th>communities and crimes</th>
</tr>
</thead>
<tbody>
<tr>
<td># of samples</td>
<td>442</td>
<td>506</td>
<td>5875</td>
<td>2195</td>
</tr>
<tr>
<td># of predictors</td>
<td>10</td>
<td>13</td>
<td>19</td>
<td>102</td>
</tr>
</tbody>
</table>

4.2 Real data examples

By applying our proposed method to several real datasets, we examined the effectiveness of our proposed procedure. We used four benchmark datasets; diabetes, Boston housing, Parkinson’s disease, and communities and crimes datasets. The diabetes dataset is available from the lars package in the software R. Remaining datasets are obtained from UCI database (http://archive.ics.uci.edu/ml/index.html). The numbers of samples and predictors for the four datasets are summarized in Table 4. Note that we deleted missing values for Parkinson’s disease and communities and crimes datasets.

We randomly and equally divided each dataset into training data and test data. Using the training data, we implemented our proposed procedures (aPIC and aPIC+SA), and then computed PSEs by the use of the test data. We repeated this procedure 200 times. In addition to our proposed procedures, we implemented DIC, DIC+SA, Blasso, Blasso+SA, and Lasso, which are introduced in Section 4.1. For all datasets, we generated 4,000 MCMC samples, and then the first 1,000 samples were discarded as burn-in. We observed that the MCMC simulations converged, since the R.hat ratios were close to one.

Figure 1 shows boxplots of the PSEs. Note that we eliminated one result for the communities and crimes dataset, since the result was clearly an outlier. From the figure, we observe that Blasso and Blasso+SA are often superior to other methods, although the two methods sometimes tend to have large variances. Meanwhile, our proposed procedures, aPIC and aPIC+SA, produce small median values of PSEs similar to Blasso and Blasso+SA, and have variances that are small and relatively stable. Therefore, we believe that aPIC and aPIC+SA may be useful in terms of yielding relatively small medians with small variances.
Figure 1: Boxplots of the PSE. (a) shows the result for the diabetes, (b) that for the Boston housing, (c) that for the Parkinson, (d) that for the communities and crimes.


5 Concluding remarks

We proposed the information criterion aPIC, which can be obtained by evaluating the Bayesian predictive distribution for Bayesian lasso regression for the selection of appropriate values of hyperparameters included in a prior distribution. For derivations of this criterion, we need to approximate the Laplace prior distribution for the coefficients. The prior distribution was approximated by the normal distribution from the viewpoints of minimizing the Kullback-Leibler information between the Laplace distribution and normal distributions. Numerical examples showed that our proposed procedure is superior to other methods in terms of prediction, estimation and model selection accuracies.

It is important to introduce information criteria by evaluating the Bayesian predictive distribution for logistic, Poisson, and Cox regressions estimated by the Bayesian lasso. In addition, it would also be interesting to derive model selection criteria for other sparse regularization methods from Bayesian viewpoints, e.g., elastic net, SCAD, and MCP by a Bayesian approach. We leave these topics as future research.

Acknowledgement

This work was supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), #24700280, 2012–2015.

References


Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy,


List of MI Preprint Series, Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The initial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in infinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the $L^2$ a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials
MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the $L_1$ regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack’s formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric $\mathcal{W}$-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications
MI2009-15  Yuya ISHIHARA & Yoshiyuki KAGEI  
Large time behavior of the semigroup on $L^p$ spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16  Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE  
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17  Masato WAKAYAMA & Keitaro YAMAMOTO  
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18  Me Me NAING & Yasuhide FUKUMOTO  
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19  Mitsunori KAYANO & Sadanori KONISHI  
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20  Shuichi KAWANO & Sadanori KONISHI  
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21  Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO  
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22  Eiji ONODERA  
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23  Stjepan LUGOMER & Yasuhide FUKUMOTO  
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24  Yu KAWAKAMI  
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25  Takehiko KINOSHITA & Mitsuhiro T. NAKAO  
On very accurate enclosure of the optimal constant in the a priori error estimates for $H^2_0$-projection

MI2009-26  Manabu YOSHIDA  
Ramification of local fields and Fontaine’s property (Pm)

MI2009-27  Yu KAWAKAMI  
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28  Masahisa TABATA  
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29  Yoshiyuki KAGEI & Yasunori MAEKAWA  
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance
MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke’s zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models

MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons

MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine

MI2010-6 Shuichi KAWANO, Toshihiro MISUMI & Sadanori KONISHI
Semi-supervised logistic discrimination via graph-based regularization

MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation

MI2010-8 Takahiro ITO
Abstract collision systems on groups
MI2010-9 Hiroshi YOSHIDA, Kinji KIMURA, Naoki YOSHIDA, Junko TANAKA & Yoshihiro MIWA
An algebraic approach to underdetermined experiments

MI2010-10 Kei HIROSE & Sadanori KONISHI
Variable selection via the grouped weighted lasso for factor analysis models

MI2010-11 Katsusuke NABESHIMA & Hiroshi YOSHIDA
Derivation of specific conditions with Comprehensive Groebner Systems

MI2010-12 Yoshiyuki KAGEI, Yu NAGAFUCHI & Takeshi SUDOU
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Poiseuille type flow

MI2010-13 Reiichiro KAWAI & Hiroki MASUDA
On simulation of tempered stable random variates

MI2010-14 Yoshiyasu OZEKI
Non-existence of certain Galois representations with a uniform tame inertia weight

MI2010-15 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of a Rotating Flow Driven by Precession of Arbitrary Frequency

MI2010-16 Yu KAWAKAMI & Daisuke NAKAJO
The value distribution of the Gauss map of improper affine spheres

MI2010-17 Kazunori YASUTAKE
On the classification of rank 2 almost Fano bundles on projective space

MI2010-18 Toshimitsu TAKAESU
Scaling limits for the system of semi-relativistic particles coupled to a scalar bose field

MI2010-19 Reiichiro KAWAI & Hiroki MASUDA
Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling

MI2010-20 Yasuhide FUKUMOTO, Makoto HIROTA & Youichi MIE
Lagrangian approach to weakly nonlinear stability of an elliptical flow

MI2010-21 Hiroki MASUDA
Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test

MI2010-22 Toshimitsu TAKAESU
A Generalized Scaling Limit and its Application to the Semi-Relativistic Particles System Coupled to a Bose Field with Removing Ultraviolet Cutoffs

MI2010-23 Takahiro ITO, Mitsuhiko FUJIO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Composition, union and division of cellular automata on groups

MI2010-24 Toshimitsu TAKAESU
A Hardy’s Uncertainty Principle Lemma in Weak Commutation Relations of Heisenberg-Lie Algebra
MI2010-25 Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators

MI2010-26 Reiichiro KAWAI & Hiroki MASUDA
On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling

MI2010-27 Chikashi ARITA & Daichi YANAGISAWA
Exclusive Queueing Process with Discrete Time

MI2010-28 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Motion and Bäcklund transformations of discrete plane curves

MI2010-29 Takanori YASUDA, Masaya YASUDA, Takeshi SHIMOYAMA & Jun KOGURE
On the Number of the Pairing-friendly Curves

MI2010-30 Chikashi ARITA & Kohei MOTEGI
Spin-spin correlation functions of the $q$-VBS state of an integer spin model

MI2010-31 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and spike detection via Gaussian basis expansions

MI2010-32 Nobutaka NAKAZONO
Hypergeometric $\tau$ functions of the $q$-Painlevé systems of type $(A_{2} + A_{1})^{(1)}$

MI2010-33 Yoshiyuki KAGEI
Global existence of solutions to the compressible Navier-Stokes equation around parallel flows

MI2010-34 Nobushige KUROKAWA, Masato WAKAYAMA & Yoshinori YAMASAKI
Milnor-Selberg zeta functions and zeta regularizations

MI2010-35 Kissani PERERA & Yoshihiro MIZOGUCHI
Laplacian energy of directed graphs and minimizing maximum outdegree algorithms

MI2010-36 Takanori YASUDA
CAP representations of inner forms of $Sp(4)$ with respect to Klingen parabolic subgroup

MI2010-37 Chikashi ARITA & Andreas SCHADSCHNEIDER
Dynamical analysis of the exclusive queueing process

MI2011-1 Yasuhide FUKUMOTO & Alexander B. SAMOKHIN
Singular electromagnetic modes in an anisotropic medium

MI2011-2 Hiroki KONDO, Shingo SAITO & Setsuo TANIGUCHI
Asymptotic tail dependence of the normal copula

MI2011-3 Takehiro HIROTsu, Hiroki KONDO, Shingo SAITO, Takuya SATO, Tatsushi TANAKA & Setsuo TANIGUCHI
Anderson-Darling test and the Malliavin calculus

MI2011-4 Hiroshi INOUE, Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling via Compressed Sensing
MI2011-5 Hiroshi INOUE
Implications in Compressed Sensing and the Restricted Isometry Property

MI2011-6 Daeju KIM & Sadanori KONISHI
Predictive information criterion for nonlinear regression model based on basis expansion methods

MI2011-7 Shohei TATEISHI, Chiaki KINJYO & Sadanori KONISHI
Group variable selection via relevance vector machine

MI2011-8 Jan BREZINA & Yoshiyuki KAGEI
Decay properties of solutions to the linearized compressible Navier-Stokes equation around time-periodic parallel flow
Group variable selection via relevance vector machine

MI2011-9 Chikashi ARITA, Arvind AYYER, Kirone MALLICK & Sylvain PROLHAC
Recursive structures in the multispecies TASEP

MI2011-10 Kazunori YASUTAKE
On projective space bundle with nef normalized tautological line bundle

MI2011-11 Hisashi ANDO, Mike HAY, Kenji KAJIWARA & Tetsu MASUDA
An explicit formula for the discrete power function associated with circle patterns of Schramm type

MI2011-12 Yoshiyuki KAGEI
Asymptotic behavior of solutions to the compressible Navier-Stokes equation around a parallel flow

MI2011-13 Vladimír CHALUPECKÝ & Adrian MUNTEAN
Semi-discrete finite difference multiscale scheme for a concrete corrosion model: approximation estimates and convergence

MI2011-14 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Explicit solutions to the semi-discrete modified KdV equation and motion of discrete plane curves

MI2011-15 Hiroshi INOUE
A generalization of restricted isometry property and applications to compressed sensing

MI2011-16 Yu KAWAKAMI
A ramification theorem for the ratio of canonical forms of flat surfaces in hyperbolic three-space

MI2011-17 Naoyuki KAMIYAMA
Matroid intersection with priority constraints

MI2012-1 Kazufumi KIMOTO & Masato WAKAYAMA
Spectrum of non-commutative harmonic oscillators and residual modular forms

MI2012-2 Hiroki MASUDA
Mighty convergence of the Gaussian quasi-likelihood random fields for ergodic Levy driven SDE observed at high frequency
MI2012-3 Hiroshi INOUE
A Weak RIP of theory of compressed sensing and LASSO

MI2012-4 Yasuhide FUKUMOTO & Youich MIE
Hamiltonian bifurcation theory for a rotating flow subject to elliptic straining field

MI2012-5 Yu KAWAKAMI
On the maximal number of exceptional values of Gauss maps for various classes of surfaces

MI2012-6 Marcio GAMEIRO, Yasuaki HIRAOKA, Shunsuke IZUMI, Miroslav KRAMAR, Konstantin MISCHAIKOW & Vidit NANDA
Topological Measurement of Protein Compressibility via Persistence Diagrams

MI2012-7 Nobutaka NAKAZONO & Seiji NISHIOKA
Solutions to a $q$-analog of Painlevé III equation of type $D_7^{(1)}$

MI2012-8 Naoyuki KAMIYAMA
A new approach to the Pareto stable matching problem

MI2012-9 Jan BREZINA & Yoshiyuki KAGEI
Spectral properties of the linearized compressible Navier-Stokes equation around time-periodic parallel flow

MI2012-10 Jan BREZINA
Asymptotic behavior of solutions to the compressible Navier-Stokes equation around a time-periodic parallel flow

MI2012-11 Daeju KIM, Shuichi KAWANO & Yoshiyuki NINOMIYA
Adaptive basis expansion via the extended fused lasso

MI2012-12 Masato WAKAYAMA
On simplicity of the lowest eigenvalue of non-commutative harmonic oscillators

MI2012-13 Masatoshi OKITA
On the convergence rates for the compressible Navier-Stokes equations with potential force

MI2013-1 Abduuwali PAERHATI & Yasuhide FUKUMOTO
A Counter-example to Thomson-Tait-Chetayev’s Theorem

MI2013-2 Yasuhide FUKUMOTO & Hirofumi SAKUMA
A unified view of topological invariants of barotropic and baroclinic fluids and their application to formal stability analysis of three-dimensional ideal gas flows

MI2013-3 Hiroki MASUDA
Asymptotics for functionals of self-normalized residuals of discretely observed stochastic processes

MI2013-4 Naoyuki KAMIYAMA
On Counting Output Patterns of Logic Circuits

MI2013-5 Hiroshi INOUE
RIPless Theory for Compressed Sensing
MI2013-6 Hiroshi INOUE
Improved bounds on Restricted isometry for compressed sensing

MI2013-7 Hidetoshi MATSUI
Variable and boundary selection for functional data via multiclass logistic regression modeling

MI2013-8 Hidetoshi MATSUI
Variable selection for varying coefficient models with the sparse regularization

MI2013-9 Naoyuki KAMIYAMA
Packing Arborescences in Acyclic Temporal Networks

MI2013-10 Masato WAKAYAMA
Equivalence between the eigenvalue problem of non-commutative harmonic oscillators and existence of holomorphic solutions of Heun’s differential equations, eigenstates degeneration, and Rabi’s model

MI2013-11 Masatoshi OKITA
Optimal decay rate for strong solutions in critical spaces to the compressible Navier-Stokes equations

MI2013-12 Shuichi KAWANO, Ibuki HOSHINA, Kazuki MATSUDA & Sadanori KONISHI
Predictive model selection criteria for Bayesian lasso