On the Behavior of MDL Denoising

Teemu Roos
Petri Myllymäki
Helsinki Institute for Information Technology
Univ. of Helsinki & Helsinki Univ. of Technology
P. O. Box 9800 FIN-02015 TKK, Finland

Henry Tirri
Nokia Research Center
P. O. Box 407 Nokia Group
FIN-00045, Finland

Abstract

We consider wavelet denoising based on minimum description length (MDL) principle. The derivation of an MDL denoising criterion proposed by Rissanen involves a renormalization whose effect on the resulting method has not been well understood so far. By inspecting the behavior of the method we obtain a characterization of its domain of applicability: good performance in the low variance regime but over-fitting in the high variance regime. We also describe unexpected behavior in the theoretical situation where the observed signal is pure noise. An interpretation for the renormalization is given which explains both the empirical and theoretical findings. For practitioners we point out two technical pitfalls and ways to avoid them. Further, we give guidelines for constructing improved MDL denoising methods.

1 INTRODUCTION

Most natural signals such as audio and images are typically redundant in that the neighboring time-slots or pixels are highly correlated. Wavelet representations of such signals are very sparse, meaning that most of the wavelet coefficients are very small and the information content is concentrated on only a small fraction of the coefficients (Mallat, 1989). This can be exploited in data compression, pattern recognition, and denoising, i.e., separating the informative part of a signal from noise. In statistics the denoising problem has been analyzed in terms of statistical risk, i.e., the expected distortion under an assumed model where typically distortion is defined as squared error and the model consists of deterministic signal plus additive Gaussian noise. Donoho & Johnstone (1994) prove that certain thresholding methods are nearly minimax optimal for a large class of signals. In the Bayesian approach a prior distribution is postulated for the signal and the expected (Bayes) risk is minimized (Ruggeri & Vidakovic, 1999). Both approaches require that parameters such as noise variance are known beforehand or determined as a part of the process.

The minimum description length (MDL) philosophy offers an alternative view where the noise is defined as the incompressible part of the signal (Rissanen, 2000). We analyze Rissanen's MDL denoising method and characterize its domain of applicability. We show that the method performs well in the low variance regime but fails in the high variance regime when compared to a thresholding method proposed by Donoho and Johnstone. In particular, in the theoretical situation where the noise completely dominates the signal, the MDL denoising method retains a majority of the wavelet coefficients even though in this case discarding all coefficients is the optimal solution in terms of both statistical risk and what we intuitively understand as separating information from noise.

We explain the behavior of the MDL method by showing that it results not from the MDL principle itself but from a renormalization technique used in deriving the method. We also point out two technical pitfalls in the implementation of MDL denoising that practitioners should keep in mind. Further, we give guidelines for constructing MDL denoising methods that have a wider domain of applicability than the current one and list objectives for future research in this direction.
2.1 STOCHASTIC COMPLEXITY

A recent introduction to MDL is given by Grünwald (2005), see also Barron et al. (1998).

We start by introducing some notation and briefly reviewing the NML (Normalized Maximum Likelihood) density for a model class parameterized by parameter vector \( \theta \). Implicit in the notation is the range of integration over the random variable \( y \) with the support \( Y \) within which the data is restricted. A range other than the full domain of \( y \) is necessary in cases where the expectation over \( y \) is restricted instead that of the the data. For these reasons the model class is restricted instead that of the the data.

For model classes with unbounded parametric complexity, is termed stochastic complexity. For model classes with unbounded parametric complexity, \( C \) is defined by

\[
C(\theta) = \int_{y \in Y} \ln \left( \frac{f_y(\theta; \hat{\theta}(y))}{\hat{C}(\theta)} \right) dy
\]

where the expectation over \( y \) is taken with respect to \( \hat{C}(\theta) \), the unique maximizer of the maximin problem with respect to \( y \).

The NML density is the unique minimizer in the set of densities \( \{ p(\theta; y) \} \) of the difference between the ideal code-length (negative of the logarithm) of the NML density and the unachievable worst-case code-length (negative of the logarithm) of the least favorable density in that set. This 'renormalized' NML can be used for model selection in linear regression and denoising. We discuss this renormalization scheme where the hyperparameters defining the range of the model class are learned from the data. Rissanen (1996) proposes to use a two-part code for a density \( q(\theta; y) \). For any indexed data \( y \), we define a minimax problem (Shtarkov, 1987):

\[
\min_{q(\theta; y)} \max_{p(\theta; y)} \ln \left( \frac{\hat{C}(\theta)}{\int_{y \in Y} \hat{C}(\theta) q(\theta; y) \, dy} \right)
\]

where \( \hat{C}(\theta) \) is a normalizing constant:

\[
\hat{C}(\theta) = \int_{y \in Y} f_y(\theta; \hat{\theta}(y)) \, dy
\]

The difference between the ideal code-length (negative of the logarithm) of the NML density and the unachievable worst-case data generating density is the (expected) regret. The minimax problem above exchanged. For these reasons the expectation over \( y \) is taken with respect to \( \hat{C}(\theta) \), the unique maximizer of the maximin problem with respect to \( y \).

2.2 PARAMETRIC COMPLEXITY

The MDL principle advocates the choice of the model class for which stochastic complexity is minimized. It is instructive to view NML as seeking a balance between fit versus complexity. The numerator measures how well the best model in the model class can represent the observed data while the denominator 'penalizes' too complex model classes. The logarithm of the denominator, \( \ln \hat{C}(\theta) \), is the problem of unbounded parametric complexity of the denominator, \( \ln \hat{C}(\theta) \), is termed stochastic complexity of the model class. Currently one of the most advantageous of the restricted range is the problem of unbounded parametric complexity of the denominator, \( \ln \hat{C}(\theta) \), is termed stochastic complexity of the model class. Currently one of the most advantageous of the restricted range. Forster & Stine (2001, 2005) analyze similar schemes where the range of the data is first encoded using a code based on an universal code for integers and then encoded using a code based on a code for integers. The renormalization and the resulting MDL denoising scheme where the hyperparameters defining the range of the observed data while the denominator 'penalizes' too complex model classes. The logarithm of the denominator, \( \ln \hat{C}(\theta) \), is the problem of unbounded parametric complexity of the denominator, \( \ln \hat{C}(\theta) \), is termed stochastic complexity of the model class. Currently one of the most advantageous of the restricted range.
zero mean and variance terms that are modeled as independent Gaussian with

\[ y = X\beta + \epsilon \]

where \( \| \cdot \| \) denotes the squared Euclidean norm. The maximum likelihood estimators of \( \beta \) and \( \sigma \) are given by

\[ \hat{\beta} = (X^T X)^{-1} X^T y \]

and

\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - X_i \hat{\beta})^2 \]

This section closely follows Rissanen (2000). Let us then obtain a regressor matrix \( X \) be square, i.e.,

\[ X = (x_{ij}) \]

where \( x_{ij} \) are arbitrary basis functions. One both theoretically and practically appealing way to define the functions is called

\[ f \]

instead of using all the basis vectors, we may also choose a subset \( \gamma \) of them. This gives the reconstruction

\[ \hat{y} = \sum_{j \in \gamma} x_{ij} \hat{\beta}_j \]


deemed version \( \hat{\gamma} \) of the original signal is left to be modeled as noise. Since the coefficients whose absolute value is less than the threshold are discarded. Using the maximum likelihood estimator of the noise variance

\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - X_i \hat{\beta})^2 \]

is seen to be the sum of the discarded coefficients divided by \( n \). We denote for convenience the squared norm of the coefficient in the full model

\[ \| \hat{\beta} \|^2 = \sum_{j=1}^{n} \hat{\beta}_j^2 \]

while the discarded coefficients would contain as much information as the retained coefficients are also shrunk towards zero. Hence, it makes it plausible to recover the informative part of the signal independent and given by

\[ X \gamma \beta + \epsilon \]

Thus, we define for each \( j \in \gamma \)

\[ \hat{y}_j = \sum_{i \in \gamma} x_{ij} \hat{\beta}_i \]

and taking as the basis functions

\[ Z = (z_{ij}) \]

where \( z_{ij} \) have

\[ \| z \| = \sqrt{\sum_{i,j} z_{ij}^2} \]

maximum likelihood estimators as the values of the retained coefficients

\[ \hat{\beta}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} y_i \]

are arbitrary basis functions. One both theoretically and practically appealing way to define the functions

\[ f \]

is called

\[ \hat{\gamma} \]

The idea of wavelet thresholding (Donoho & Johnstone, 1991; Mallat, 1989) makes it plausible to recover the informative part of the signal.
to assume a hypothetical generating model whose ex-

One of the most characteristic features of the MDL ap-

4.1 MDL APPROACH TO DENOISING

for describing the signal.

respectively, thus depending on the model class used

compressible and the incompressible part of the signal

and noise in the observed signal. Unlike in the statisti-

The MDL principle offers a different approach to de-

4 MDL DENOISING

enen

Johnstone method and MDL denoising.

methods are not discussed in the current paper. Fodor

limitations and the fact that our real focus is in under-

by the mentioned authors and others but due to space

eral other, more refined denoising methods suggested

mally works well as long as the signal is contained mainly

median of the coeffi-

Donoho & Johnstone sug-

recommend using as an estimator the median of the coeffi-

In order to apply the method in practice, one usu-

σ

where

In statistical wavelet denoising the denoising problem

In denoising, MDL model selection is performed by

expected (Bayes) risk is minimized. Donoho & John-

eri & Vidakovic, 1999; Chang

maximal value

equality on the true signal is postulated and the

This requires an assumed model typically involv-

nal. This requires an assumed model typically involv-

plane)

In fact even the renormalization requires the data

range to be restricted but it turns out that the final range

range is first restricted such that the squared (Eu-

data range is first restricted such that the squared (Eu-

bounded and NML is not defined unless the range of

based models and more generally, for linear regres-

considering each subset of the coefficients as a model

is clearly much easier to accept than the assumption

that the assumed model is indeed an exact replica of

existence would be very hard to verify. Any background

incorporated in the choice of the model class. The only

information regarding the phenomenon under study is

that at least one of the model classes un-

After the application of Stirling's approximation to

ance ˆ

R

estimated two-part and mixture codes for wavelet denois-

havior of MDL based denoising, these

nem

don't affect the resulting criterion.

range to be restricted but it turns out that the final range

is incorrect in some of the earlier publications.

Multiplying the code-length formula by two gives an

S

k

γ

y

n

σ

√

n,

γ

β

∥

γ

y

n

k

(6)

(7)

k

2

ln

k

(−

n

2

−

n

k

2

ln(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2

ln

k

(−

n

k

2
It can be shown that the criterion is always maximized by choosing $\gamma$ such that either the $k$ largest or the $k$ smallest coefficients are retained for some $k$. We consider this an artefact of the renormalization performed and assume in the what follows that the $k$ largest coefficients are retained. We return to the issue in Sec. 5.3.

4.3 PRACTICAL ISSUES

We point out two issues of a rather technical nature that nevertheless deserve to be noted by practitioners since we have found them to result in very poor performance in more than one case. First, in all wavelet thresholding methods, it should be made sure that the wavelet transform used is such that the coefficients are scaled properly, in other words, that the corresponding basis is orthogonal. This is essential for all wavelet thresholding methods. It is easy to check that the sum of squares of the original data and the transformed coefficients are always equal.

Secondly, since the criterion is derived for continuous data and involves densities, problems may occur when it is applied to low-precision or discrete, say integer, data. If the data can be represented exactly by some number $k_0$ of coefficients, the criterion becomes minus infinity for all $k \geq k_0$ because the first term includes a logarithm of zero. Also, for $k$ almost as large as $k_0$ the criterion takes a very small value and such a value of $k$ is often selected as the optimal one potentially resulting in severe over-fitting. This problem may either be solved by using a lower bound for $(S - S_\gamma) / (n - k)$ corresponding to a lower bound on the variance. Alternatively, once a sudden drop to minus infinity in the criterion is recognized it is possible to reject all values of $k$ that are near the point where the drop occurs.

5 BEHAVIOR OF MDL DENOISING

By inspecting the behavior of the MDL denoising criterion as a function of noise variance, we are able to give a rough characterization of its domain of applicability. This makes way towards a more important goal, the understanding of renormalized NML, and potential ways of generalizing and improving it.

5.1 EMPIRICAL OBSERVATIONS

Fig. 1 illustrates the behavior of the MDL denoising method and the method by Donoho & Johnstone described in Sec. 3 with Daubechies N=4 wavelet basis. The original image is distorted by Gaussian noise to get a noisy signal. When there is little noise, the difference is small, MDL method performing better in terms of standard error. However, when there is much noise the methods produce very different results. The Donoho-Johnstone method retains only 0.3 percent of the coefficients while the MDL method retains 46.9 percent of them, the former giving a better result in terms of standard error.

The effect of the standard deviation of noise on the behavior of the two methods can be clearly seen in Fig. 2. It can be seen that the MDL method outperforms the Donoho-Johnstone method when the noise standard deviation is less than 15. However, outside this range the performance of the MDL method degrades linearly...
Essentially, we need to evaluate the asymptotics of Gaussian noise. Since the criterion is scale invariant we may without loss of generality assume unit variance.

The degradation of performance of the MDL denoising criterion is underlined when the noise variance is very large. This can be demonstrated theoretically by considering what happens when the noise variance grows without bound so that in the limit the signal is pure Gaussian noise.

Figure 2: Effect of noise.

As a rough characterization of the domain of applicability of the MDL method it can be said that the standard deviation of the signal determines the standard error of the noise should be compared to the standard deviation of the original signal which in this case was due to retaining too many coefficients. The standard deviation of the signal determines the statistical risk and the natural meaning of information.

Figure 3: The renormalized NML denoising criterion with sample sizes are \( n=128 \) (on the left), and \( n=1024 \) (on the right), with 50 repetitions.

Now in order to contain a fraction of Gaussian noise model is equal to one, the expectation becomes

\[
E\left[\sum_{i=1}^{n} x_i^2\right] = \sum_{i=1}^{n} E(x_i^2) = \sum_{i=1}^{n} \frac{1}{2}\pi \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} n
\]

The theoretical minima for the two samples obtained once the expectation of the squared coefficients can be easily calculated. This is suboptimal in terms of both statistical risk and the natural meaning of information.

\[
\text{asymptotic curve is plotted with a solid line. By evaluating the expectation of all coefficients under the unit variance, the largest coefficients in absolute value. Let } \beta_k \text{ the squared sum of the } k \text{ largest coefficients in ascending order. We have that the first retained coefficient is unique. If we consider a fixed parameter instead of a random variable, the terms in the above sum are independent with expectation given by:}
\]

\[
E\left[\sum_{i=1}^{n} |x_i|^2 \mid t\right] = 1
\]

\[
E\left[\sum_{i=1}^{n} x_i^2 \mid t\right] = \sum_{i=1}^{n} E(x_i^2) = \sum_{i=1}^{n} \frac{1}{2}\pi \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} n
\]

The integral is given by

\[
\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}
\]

Now in order to contain a constant fraction of Gaussian noise model is equal to one, the expectation becomes

\[
E\left[\sum_{i=1}^{n} x_i^2\right] = \sum_{i=1}^{n} E(x_i^2) = \sum_{i=1}^{n} \frac{1}{2}\pi \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} n
\]

Theoretical minima for the two samples are \( k=32 \) respectively. The asymptotic behavior of the average standard deviation is known.

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i^2 = \frac{1}{2}\pi \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2}
\]

Now in order to contain a constant fraction of Gaussian noise model is equal to one, the expectation becomes

\[
E\left[\sum_{i=1}^{n} x_i^2\right] = \sum_{i=1}^{n} E(x_i^2) = \sum_{i=1}^{n} \frac{1}{2}\pi \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} n
\]

The integral is given by

\[
\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}
\]

Now in order to contain a constant fraction of Gaussian noise model is equal to one, the expectation becomes

\[
E\left[\sum_{i=1}^{n} x_i^2\right] = \sum_{i=1}^{n} E(x_i^2) = \sum_{i=1}^{n} \frac{1}{2}\pi \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2\pi}}{2} n
\]

The integral is given by

\[
\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}
\]
which is seen to be proportional to $\|\gamma\|$, where $\gamma$ denotes the complement of $\gamma$. 

The density of the following form (ignoring normalization)

\[
\text{the code-length function (7) is the negative logarithm of a corresponding}
\]

the above described behavior. The code-length function $NML$ denoising criterion in order to understand

Let us now consider the interpretation of the renormalization points of two Gaussian densities fitted to the discarded coefficients.

As long as the variance of the informative coefficients, as well with variance equal to the sum of the signal variance is large compared to the noise variance, the criterion performs well. From orthogonality of the retained and discarded coefficients.

It has been observed that wavelet coefficients in natural images tend to be well modeled by general

similar reasoning also shows that while the criterion is symmetric in the two sets of coefficients, one should always consider Bayes optimal hard thresholding in this model

Ruggeri & Vidakovic (1999) estimate the scale parameters $\sigma_k$ from the observed signal. The construction of an $NML$ model based on Laplacian and generalized Gaussian densities is an interesting future research topic.

Figure 4: Gaussian densities fitted to noisy Lena ($\sigma_0$). The empirical histogram is plotted with solid line.

$\beta_k$ is a normalization constant (Mallat, 1989). The typical values of $\beta_k$ where $\beta_k = 153$.

Figure 4. As long as the noise variance is not too high.

$\sigma_2$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_2$ is where $\beta_2 = 6$. As long as the noise variance is not too high.

It is quite easy to derive rough conditions on when

$\beta_3$ is formed Gaussian densities with variance adjusted for the discarded coefficients.

The typical values of $\beta_3$ where $\beta_3 = 4$. It is also easy

$\sigma_4$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_4$ is where $\beta_4 = 15$. It is quite easy to derive rough conditions on when

$\beta_5$ is formed Gaussian densities with variance adjusted for the discarded coefficients.

The typical values of $\beta_5$ where $\beta_5 = 7$. It is also easy

$\beta_6$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_6$ is where $\beta_6 = 6$. As long as the noise variance is not too high.

$\beta_7$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_7$ is where $\beta_7 = 4$. As long as the noise variance is not too high.

$\beta_8$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_8$ is where $\beta_8 = 153$. As long as the noise variance is not too high.

$\beta_9$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_9$ is where $\beta_9 = 153$. As long as the noise variance is not too high.

$\beta_{10}$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_{10}$ is where $\beta_{10} = 153$. As long as the noise variance is not too high.

$\beta_{11}$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_{11}$ is where $\beta_{11} = 153$. As long as the noise variance is not too high.

$\beta_{12}$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_{12}$ is where $\beta_{12} = 153$. As long as the noise variance is not too high.

$\beta_{13}$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_{13}$ is where $\beta_{13} = 153$. As long as the noise variance is not too high.

$\beta_{14}$ is near one which corresponds significantly larger than that of the noise coefficients.

$\beta_{14}$ is where $\beta_{14} = 153$. As long as the noise variance is not too high.


References

CAL Network of Excellence, IST-2002-506778. This programme of the European Community, under the PAS-to PAS-project, the Finnish Technology Agency under project PMMA, and the IST Projects Minos and Cepler, the Finnish Technology Agency.

sula Gather for useful discussions. This work was supported in part by the Academy of Finland and is of considerable interest not only in denoising applications but in the MDL framework in general.

We thank Jorma Rissanen, Peter Grünwald, and Uwe Gather for useful discussions. This work was supported in part by the Academy of Finland and is of considerable interest not only in denoising applications but in the MDL framework in general.

Acknowledgments

We discussed a solution by Rissanen involving a renormalization whose effect has been unclear so far. We gave an interpretation of this renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

The interpretation also facilitates understanding of the renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

The reported empirical and theoretical findings suggested a characterization of the domain of applicability for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

6 CONCLUSIONS

In its general form, the MDL principle essentially aims at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

There are, however, some intricate issues to avoid them. We gave an interpretation of the renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

The interpretation also facilitates understanding of the renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

We gave an interpretation of the renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

The reported empirical and theoretical findings suggested a characterization of the domain of applicability for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

There are, however, some intricate issues to avoid them. We gave an interpretation of the renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

The reported empirical and theoretical findings suggested a characterization of the domain of applicability for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

There are, however, some intricate issues to avoid them. We gave an interpretation of the renormalization by showing that it results in a possibility for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.

The reported empirical and theoretical findings suggested a characterization of the domain of applicability for Rissanen's denoising method. It was seen at separating meaningful information from noise, and is of considerable interest not only in denoising applications but in the MDL framework in general.