Hyperparameter estimation in image restoration

2008 J. Phys. A: Math. Theor. 41 332004

(http://iopscience.iop.org/1751-8121/41/33/332004)

IOP FTC

IOP Publishing — Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **41** (2008) 332004 (7pp) doi:10.1088/1751-8113/41/33/332004

FAST TRACK COMMUNICATION

Hyperparameter estimation in image restoration

Hirohito Kiwata

Division of Natural Science, Osaka Kyoiku University, Kashiwara, Osaka 582-8582, Japan

E-mail: kiwata@cc.osaka-kyoiku.ac.jp

Received 8 May 2008, in final form 1 July 2008
Published 18 July 2008
Online at stacks.iop.org/JPhysA/41/332004

Abstract
The hyperparameter in image restoration by the Bayes formula is an important quantity. This communication shows a physical method for the estimation of the hyperparameter without approximation. For artificially generated images by prior probability, the hyperparameter is computed accurately. For practical images, accuracy of the estimated hyperparameter depends on the magnetization and energy of the images. We discuss the validity of prior probability for an original image.

PACS numbers: 02.50.−r, 05.50.+q, 07.05.Pj, 95.75.Mn

Mathematical methods in statistical physics have been applied to information processing problems [1]. A probabilistic model has been used to construct the problems. An analogy between a probabilistic model and a formula of statistical physics supports the validity for applicability. One of the main topics which deals with statistical-physics approaches to information processing problems is image restoration [2]. An image on a computer is represented by a sequence of bits. When a digital image is transferred through a channel, it is corrupted by noise. The purpose of image restoration is to restore an original image from a degraded image.

The problem is to infer an original image from a degraded image. The Bayes formula plays an important role in the image restoration by a probabilistic method [3]. The Bayes formula is expressed by (posterior prob.) $\propto$ (conditional prob.) $*$ (prior prob.). The conditional probability and the prior probability contain parameters (hyperparameters). In order to obtain a properly restored image by the posterior probability of the Bayes formula, it is necessary to use appropriate values of the hyperparameters [4]. However, one usually has only a degraded image and no knowledge of a degradation process characterized by the conditional probability and an original image characterized by the prior probability. For the sake of simplicity, we assume that the conditional probability is given by a memoryless symmetric channel. We have to estimate the hyperparameters from a degraded image [5–9]. So far the hyperparameters are determined by maximizing a marginal likelihood function (MML) [10]. However, a computational task for the summation in marginalization is exponentially huge. One has to resort to simulation or approximate methods (mean-field or Bethe approximation) to implement the method. In the present communication, we demonstrate a physical method for obtaining the hyperparameters without any approximations. Morita and Tanaka attempted to estimate the hyperparameters on an additional assumption about images [11]. By comparison with their method, our method is simple and natural.

We consider a binary (black and white) image. A black pixel represents 0 as bit expression or down spin in the Ising model. A white pixel represents 1 as bit expression or up spin in the Ising model. When an original image is corrupted by noise, one receives a degraded image with the state of a pixel inverted from an original value with probability p. In a binary symmetric channel a change from the state of each pixel to another state occurs with the same probability p independently of the other pixels. The probability p is one of the hyperparameters. We examine a change of the number of black or white pixels between an original image and a degraded image:

$$\begin{cases} N_1' = (1-p)N_1 + pN_0 \\ N_0' = pN_1 + (1-p)N_0, \end{cases} \tag{1}$$

where N_1 and N_0 are the number of white and black pixels in an original image, respectively. The prime means the number of pixels in a degraded image. Solving the eigenvalue problem, we obtain the following relations:

$$N_1' + N_0' = N_1 + N_0, \tag{2}$$

$$N_1' - N_0' = (1-2p)(N_1 - N_0). \tag{3}$$

Equation (2) shows conservation of the total number of pixels between an original image and a degraded image. From the point of view of Ising spin, Equation (3) means a change of magnetization, which is defined by the difference between the number of sites with up spin and that with down spin. Equation (3) can be expressed in the following way,

$$M' = (1-2p)M, \tag{4}$$

where M and M' are the magnetization of an original image and a degraded image, respectively.

We consider a change of the number of neighboring-pixel pairs between an original image and a degraded image:

$$\begin{cases} N_{11}' = (1-p)^2 N_{11} + p(1-p)N_{10} + p^2 N_{00} \\ N_{10}' = 2p(1-p)N_{11} + [(1-p)^2 + p^2]N_{10} + 2p(1-p)N_{00} \\ N_{00}' = p^2 N_{11} + p(1-p)N_{10} + (1-p)^2 N_{00}, \end{cases} \tag{5}$$

where N_{11} is the number of neighboring-pixel pairs with 1 as a bit on both ends. Similarly N_{10} and N_{00} are defined. The prime means those in a degraded image. Solving the eigenvalue problem, we obtain the following relations,

$$N_{11}' + N_{10}' + N_{00}' = N_{11} + N_{10} + N_{00}, \tag{6}$$

$$N_{11}' - N_{00}' = (1-2p)(N_{11} - N_{00}), \tag{7}$$

$$N_{11}' - N_{10}' + N_{00}' = (1-2p)^2(N_{11} - N_{10} + N_{00}). \tag{8}$$

The first expression (6) shows conservation of the total number of neighboring pairs between an original image and a degraded image. The formula $(N_{11} - N_{10} + N_{00})$ in the third expression (8) is equal to the minus energy of the Ising model, $-H = \sum_{\langle i,j \rangle} \sigma_i \sigma_j$. Equation (8) can be expressed in the following way,

$$E' = (1-2p)^2 E, \tag{9}$$

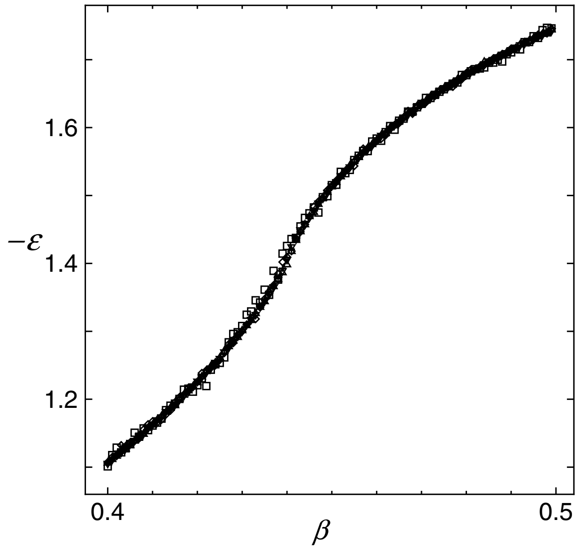

Figure 1. Energy per site as a function of β for several system sizes, 64×64, 128×128, 256×256, 512×512 and 1024×1024. The exact energy of the Onsager solution accords with the curve.

where E and E' are the energy of an original image and a degraded image respectively. Equations (4) and (9) are key concepts in the present paper. Eliminating the quantity $(1 - 2p)$, we can derive the relation

$$\frac{(M')^2}{E'} = \frac{(M)^2}{E}. \tag{10}$$

The quantity $(M)^2/E$ is constant between an original image and a degraded image. Taking the derivation process into consideration, we conclude that equation (10) bears no relation to the Bayes formula as well as the prior probability. The left-hand side of (10) can be calculated from a degraded image. If we assume that the prior probability is the Gibbs distribution of the Ising model, we can obtain β (inverse temperature) dependence of the magnetization and the energy. The quantity β characterizes the prior probability and is one of the hyperparameters. The value of β can be evaluated from relation (10) by adjusting β to equalize both sides of (10). The probability p is derived through relation (4) by

$$p = \frac{1}{2}\left[1 - \frac{M'}{M(\beta)}\right]. \tag{11}$$

In order to investigate β dependence of the magnetization and the energy of the Ising model, we perform Monte Carlo simulation. Figures 1 and 2 show results of the Monte Carlo simulation for the two-dimensional Ising model. Figure 1 indicates β dependence of the energy for several system sizes. The energy curve is independent of the system size. Although the exact solution of the Ising model is derived on an infinite two-dimensional square lattice, we use the exact solution for $E(\beta)$. Figure 2 indicates β dependence of the magnetization for several system sizes. The magnetization as a function of β depends on the system size below the critical value of β. In order to determine the β dependence of the magnetization, we use a finite-size scaling relation. The inset in figure 2 shows a finite-size scaling function of the magnetization, scaled temperature versus scaled magnetization. We derive the scaling

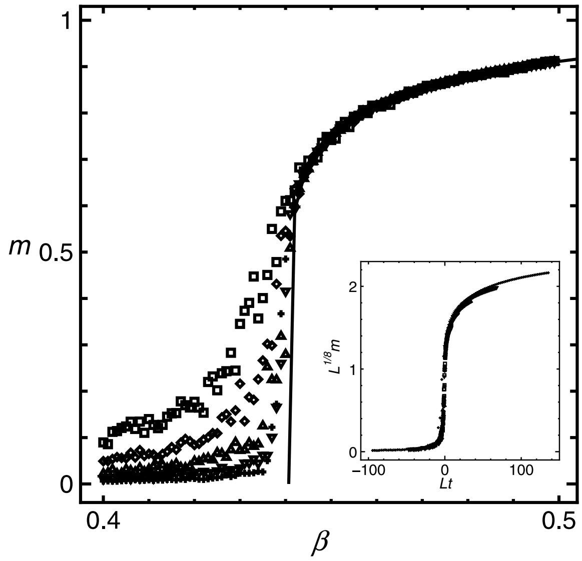

Figure 2. Magnetization per site as a function of β for several system sizes, which are the same as figure 1. The solid curve indicates the exact magnetization of the Onsager solution. The inset shows a finite-size scaling function of the magnetization. The scaling function uses exact scaling parameters.

Table 1. Average of estimated hyperparameters for 50 original images which are generated by Monte Carlo simulation on the Gibbs distribution of the Ising model.

Prior prob.	Conditional prob.	MML with mean-field approx.	MML with Bethe approx.	Our method
$\beta = 0.43$	$p = 0.1$	$p = 0.026\,106$	$p = 0.078\,867$	$p = 0.108\,118$
		$\beta = 0.305\,503$	$\beta = 0.384\,960$	$\beta = 0.404\,151$
$\beta = 0.43$	$p = 0.2$	$p = 0.076\,367$	$p = 0.166\,563$	$p = 0.205\,954$
		$\beta = 0.288\,261$	$\beta = 0.377\,584$	$\beta = 0.405\,708$
$\beta = 0.44$	$p = 0.1$	$p = 0.028\,072$	$p = 0.081\,310$	$p = 0.099\,378$
		$\beta = 0.314\,197$	$\beta = 0.398\,500$	$\beta = 0.440\,023$
$\beta = 0.44$	$p = 0.2$	$p = 0.077\,628$	$p = 0.168\,992$	$p = 0.199\,787$
		$\beta = 0.292\,880$	$\beta = 0.388\,504$	$\beta = 0.440\,021$

function from the data of figure 2 by curve fitting. By the obtained scaling function, the magnetization for any system size can be obtained as a function of β.

In order to examine the validity of our method, we implement our method for artificially created images. The original images (256×256 pixels) are generated by Monte Carlo simulation with the prior probability which is the Gibbs distribution of the Ising model with $\beta = 0.43$ and $\beta = 0.44$. The choice of the above β stems from the fact that for smaller β value than the selected values images with disordered patterns are generated and for larger β value than the selected values almost black or white images are generated by the prior probability. The original images are degraded by the binary symmetric noise with $p = 0.1$ and $p = 0.2$. Table 1 shows estimated values of the hyperparameters by our method from the degraded images. Results of MML with the mean-field approximation or the Bethe approximation are

Figure 3. Actual binary images (256 × 256 pixels), 'home' (left) and 'mandrill' (right).

Table 2. Estimated hyperparameters for actual binary images.

Image	Conditional prob.	Estimated hyperparameter
Home	$p = 0.1$	$p = -0.042\,946$
		$\beta = 0.366\,966$
Home	$p = 0.2$	$p = 0.091\,239$
		$\beta = 0.366\,677$
Mandrill	$p = 0.1$	$p = 0.105\,153$
		$\beta = 0.436\,895$
Mandrill	$p = 0.2$	$p = 0.203\,478$
		$\beta = 0.436\,792$

also listed for comparison. Our method provides accurate values of the hyperparameters. When $\beta = 0.43$, the magnetization becomes a very small quantity and so it is difficult to improve accuracy of the calculations.

We also implement our method for practical images exemplified in figure 3, 'home' and 'mandrill'. The two binary images are obtained from the standard images by using threshold processing. We produce degraded images from the original images 'home' and 'mandrill' by the degradation process. Table 2 shows the hyperparameters evaluated by our method from the degraded images. While the estimated hyperparameters of the image 'mandrill' are appropriate values, those of the image 'home' are inappropriate. The inappropriate values result from the fact that the original image 'home' is never generated by the Gibbs distribution of the Ising model. In order to investigate the validity of the prior probability, we compare β_M estimated from the magnetization of the original image with β_E estimated from the energy. The inverse temperatures estimated from the magnetization and the energy of the original image 'home' are $\beta_M = 0.380\,819$ and $\beta_E = 0.500\,947$ respectively. The inverse temperatures estimated from the magnetization and the energy of the original image 'mandrill' are $\beta_M = 0.436\,714$ and $\beta_E = 0.432\,313$ respectively. The validity of the prior probability depends on whether β_M estimated from the magnetization is close to β_E estimated from the energy. In comparison with images generated by the Gibbs distribution of the Ising model, the original image 'home' has large clusters of neighboring pixels having the same state.

The framework of the binary system is also applicable to restoration of q-valued images. Each pixel takes one of the gray-levels $(0, 1, 2, \ldots, q-1)$, where 0 and $q-1$ correspond to black and white respectively. We assume that a degradation process is given by the symmetric channel. The degradation process changes intensity of each pixel to other intensity with

Table 3. Estimated hyperparameters for 4-valued images (256 × 256 pixels).

Conditional prob.	Artificial image ($\beta = 1.09$)	Home	Mandrill	Lenna
$p = 0.05$	0.051 132	0.035 825	0.078 340	0.025 158
$p = 0.1$	0.100 191	0.090 005	0.121 588	0.081 249
β_M	1.089 335	1.104 276	1.125 190	1.098 326
β_E	1.087 304	1.135 939	1.098 612	1.150 626

probability $(q-1)p$. Through derivation similar to the binary system, the conserved quantity between an original image and a degraded image is given by

$$\frac{(M')^2}{qE' - 2N} = \frac{(M)^2}{qE - 2N}, \tag{12}$$

where N is the total number of pixels. Although the quantities E and M correspond to the energy and the magnetization respectively, they are not naive extension of the Ising model. Equation (12) also bears no relation to prior probability. If we set prior probability, β dependence of E and M is determined. In the case that prior probability is the Gibbs distribution of the Potts model [12], we have to settle β dependence of E and M by Monte Carlo simulation. As for the β dependence of M, we use a finite-size scaling relation. The left-hand side of (12) can be derived from a degraded image. The hyperparameter β is so adjusted for the right-hand side of (12) to coincide with the left-hand side. After we obtain a value of β, the conditional probability p is derived from the following relation:

$$p = \frac{1}{q}\left[1 - \frac{M'}{M(\beta)}\right]. \tag{13}$$

We implement our method for 4-valued standard images, 'home', 'mandrill' and 'lenna'. Table 3 shows estimated hyperparameters from degraded images. The result for an artificially generated image by Monte Carlo simulation with the Gibbs distribution of the Potts model is listed for comparison. The artificial image is chosen at random. For the artificially generated image the hyperparameter is estimated by our method satisfactorily. For the practical 4-valued images the results of our method are not satisfactory. Our assumption that the original images are generated by the prior probability is inappropriate for the practical images. In the case of the practical images, estimated β_M widely differs from β_E. This means that the practical images are never generated from the prior probability.

We showed the method for estimating the hyperparameters in the image restoration. The formula bears no relation to prior probability, and so our method is applicable to other prior probability [13]. So far the hyperparameters are estimated by maximization of a marginal likelihood function. The summation in marginalization is an exponentially increasing task. Instead of the summation, the mean-field approximation or the Bethe approximation has been adopted. In our method we use the exact solution or the results of Monte Carlo simulation in place of the summation. A digital image is represented by a finite number of pixels. Our method takes into consideration the size of an image. The maximization of a marginal likelihood function with the approximate methods takes no account of the finiteness of an image size. Since our method does not rely on any approximations, we can investigate appropriateness of the prior probability. In order to obtain proper values of the hyperparameters, the magnetization and the energy of the original image have to be close to those evaluated by the prior probability. The validity of the assumption for the prior probability is assessed by the energy and the magnetization of the original image. If we find prior probability which generates an image, the hyperparameters are estimated by our method accurately.

References

[1] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford: Oxford University Press)
[2] Tanaka K 2002 *J. Phys. A: Math. Gen.* **35** R81
[3] Marroquin J, Mitter S and Poggio T 1987 *J. Am. Stat. Assoc.* **82** 76
[4] Nishimori H and Michael Wong K Y 1999 *Phys. Rev.* E **60** 132
[5] Besag J 1986 *J. R. Statist. Soc.* B **48** 259
[6] Lakshmanan S and Derin H 1989 *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 799
[7] Zhang J 1992 *IEEE Trans. Signal Process.* **40** 2570
[8] Zhang J, Modestino J W and Langan D A 1994 *IEEE Trans. Image Process.* **3** 404
[9] Zhou Z, Leahy R M and Qi J 1997 *IEEE Trans. Image Process.* **6** 844
[10] Pryce J M and Bruce A D 1995 *J. Phys. A: Math. Gen.* **28** 511
[11] Morita T and Tanaka K 1996 *Physica* A **223** 244
[12] Wu F Y 1982 *Rev. Mod. Phys.* **54** 235
[13] Tanaka K, Inoue J and Titterington D M 2003 *J. Phys. A: Math. Gen.* **36** 11023