Sublinear Computational Time Modeling by Momentum-Space Renormalization Group Theory in Statistical Machine Learning Procedures

Kazuyuki Tanaka¹ · Masayuki Ohzeki^{1,2} · Muneki Yasuda³

Received: 13 February 2019 / Accepted: 22 August 2019 / Published online: 7 September 2019
© The Author(s) 2019

Abstract

We review sublinear computational time modeling using momentum-space renormalization group approaches in the statistical machine learning algorithms. The modeling scheme has been proposed and the basic frameworks have been briefly explained in a short note (Tanaka et al. in J. Phys. Soc. Jpn, 87(8), Article ID: 085001, 2018). We present their detailed formulations and some numerical experimental results of sublinear computational time modeling based on the momentum-space renormalization scheme.

Keywords Statistical machine learning · Bayesian modeling · Probabilistic graphical model · Renormalization group · Statistical mechanical informatics

1 Introduction

Probabilistic graphical models have useful applications in statistical machine learning theory, and many authors are investigating their applications in artificial intelligent computational systems [1, 2]. Realizations of statistical machine learning systems with probabilistic graphical models entail massive computational time. For avoiding such problems, some statistical–mechanical techniques have been

✉ Kazuyuki Tanaka
kazu@tohoku.ac.jp

Masayuki Ohzeki
mohzeki@tohoku.ac.jp

Muneki Yasuda
muneki@yz.yamagata-u.ac.jp

¹ Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan

² Institute of Inovative Research, Tokyo Institute of Technology, Yokohama 226-8503, Japan

³ Graduate School of Science and Engineering, Yamagata University, Yamagata pref., Yonezawa 992-8510, Japan

introduced, for example, advanced mean-field methods including belief propagations [3–8]. However, it is very difficult to reduce the computational time to the sublinear time for each system size in such advanced mean-field methods although they can achieve the linear computational time modeling. Recent research in data sciences has helped understand the necessity to treat high-dimensional state vectors, and it is necessary to achieve novel sublinear computational time modeling schemes in probabilistic graphical models. Sublinear computational time modeling is used to reduce the computational time for solving the given problem to the order of sublinear time with respect to the system size.

The renormalization group (RG) theory [9–13] is one of important theories in statistical mechanics. The RG theory was introduced to investigate a universality class of phase transitions in probabilistic graphical models by considering spatially coarse-graining procedures. The RG scheme can be divided to two kinds: real-space renormalization group (RSRG) schemes and momentum-space renormalization group (MSRG) schemes. MSRG schemes are implemented in the spectrum space of the Fourier transformations in the probabilistic graphical model and have produced many successful results in combination with perturbation techniques.

Recently, the present authors have proposed applications of RG techniques to achieve sublinear computational time modeling schemes in the statistical machine learning theory. In physics, it is well known that the RG scheme is typically used to capture critical phenomena characterized by certain diverging statistical quantities, particularly, the correlation length. It might seem that the achievements of the sublinear time computational time modeling in the statistical machine learning theory are far from the standard physical applications of the RG schemes, because the statistical models for data sciences have finite sizes and no critical phenomena are expected. However, it has become necessary to treat many high-dimensional state vectors in recent data sciences and it is important to achieve to reduce the computational time in the statistical models containing such high-dimensional state vectors. The present authors have considered introducing RG techniques to systematically reduce the system size in the statistical model while applying the statistical machine learning schemes, systematically. One way to accomplish them is to apply the RSRG scheme to probabilistic image segmentations, in which a Potts model is employed as a prior probability [14]. The other one is to introduce the momentum-space renormalization scheme to probabilistic noise reductions, in which a Gaussian graphical model is introduced as a prior probability [15]. The most important aspect of the statistical machine learning algorithms is how the computational time of the estimation of hyperparameters from data vectors should be reduced. One of familiar and practical estimation schemes involving hyperparameters is the expectation-maximization (EM) algorithm [16, 17]. However, the EM algorithm needs some commutations of statistical quantities in the massive probabilistic graphical model. Two of the works by the present authors applied the RG approaches to estimate the hyperparameters by means of the EM algorithm.

The present paper is not an original paper but a review paper of the sublinear computational time modeling using the MSRG approaches in the EM algorithm for statistical machine learning system based on the Gaussian graphical models which has been published as a short note paper [15] and their background formulations. In the present

paper, we provide their detailed formulations and some numerical experimental results of the sublinear computational time modeling based on the momentum-space renormalization scheme. Their essential part of the sublinear computational time modeling scheme has been proposed and the most important numerical experimental results have been shown as the above short note paper. In Sect. 2, we present the sublinear computational time modeling of statistical machine learning systems for probabilistic noise reductions by introducing the momentum-space renormalization schemes. Section 3 shows practical schemes of the sublinear computational time EM algorithm. In Sect. 4, statistical performance analysis of the hyperparameter estimations using the sublinear computational time EM algorithm is given. In Sect. 5, we mention some concluding remarks.

2 Momentum-Space Renormalization Group Scheme of Gaussian Graphical Model

In this section, we review fundamental formulations of momentum-space renormalization group approaches in Gaussian graphical models. Gaussian graphical models are sometimes referred also to as Gaussian–Markov random fields [4, 18–24], and are familiar probabilistic graphical models for Bayesian noise reduction systems with the EM algorithm. In noise reductions of image processing in particular, the graphical structures are usually square grid graphs and then unitary matrices based on discrete Fourier transformations are useful. Although scale transformations in the MSRG approaches are achieved through the wave-number space in the discrete Fourier transformations, we mention that they differ from low-pass filters in conventional signal processing.

Now we consider images on the square grid graph with the periodic boundary conditions for the abscissa and the ordinate. In the position vector for each pixel of the present square grid graph, the abscissa and the ordinate are denoted by x and y , respectively, such that the position vector of the pixel is denoted by (x, y) . The periodic boundary conditions for the abscissa and ordinate mean that $x = M$ and $y = N$ are interpreted as $x = 0$ and $y = 0$, respectively. We introduce the set of all the pixels by a notation $V(M, N) = \{(x, y) | x = 0, 1, \dots, M - 1; y = 0, 1, \dots, N - 1\}$. We remark that $V(M, N)$ can be regarded as an ordered set by considering the two-dimensional position vector (x, y) as

$$\mathbf{r}_i \equiv \left((i - 1) \bmod(MN), \left\lfloor \frac{i - 1}{M} \right\rfloor \right) \quad (i = 1, 2, 3, \dots, MN), \quad (1)$$

where the notation $\lfloor a \rfloor$ is the floor function defined by $\lfloor a \rfloor \equiv a - a \bmod 1$.

We define the state variables $f_{x,y}$ and $g_{x,y}$ ($(x, y) \in V(M, N)$) of a light intensity at each pixel (x, y) , in the original images and the degraded images. The state variable $f_{x,y}$ and $g_{x,y}$ take any real values in the interval $(-\infty, +\infty)$. The state vector $\mathbf{f}$ and $\mathbf{g}$ are defined by $\mathbf{f} \equiv (f_{x,y} | (x, y) \in V(M, N))$ and $\mathbf{g} \equiv (g_{x,y} | (x, y) \in V(M, N))$, respectively.

We consider a prior probability density function $\rho^{(M,N)}(f|\alpha, \gamma)$ defined by

$$\rho^{(M,N)}(f|\alpha, \gamma) \equiv \frac{1}{Z_{M,N}(\alpha, \gamma)} \exp \left(-\frac{1}{2} \alpha \sum_{\{x,y\} \in V(M,N)} \left((f_{x,y} - f_{x+1,y})^2 + (f_{x,y} - f_{x,y+1})^2 \right) - \frac{1}{2} \gamma \sum_{(x,y) \in V(M,N)} f_{x,y}^2 \right), \quad (2)$$

for original images f . Here $Z_{M,N}(\alpha, \gamma)$ is a normalization constant of $\rho^{(M,N)}(f|\alpha, \gamma)$, which is defined by

$$Z_{M,N}(\alpha, \gamma) \equiv \left(\prod_{(x,y) \in V(M,N)} \int_{-\infty}^{+\infty} dz_{x,y} \right) \times \exp \left(-\frac{1}{2} \alpha \sum_{\{x,y\} \in V(M,N)} \left((z_{x,y} - z_{x+1,y})^2 + (z_{x,y} - z_{x,y+1})^2 \right) - \frac{1}{2} \gamma \sum_{(x,y) \in V(M,N)} z_{x,y}^2 \right). \quad (3)$$

Degraded images g are generated from a given original image f according to the following conditional probability density function:

$$\rho^{(M,N)}(g|f, \beta) = \sqrt{\left(\frac{\beta}{2\pi} \right)^{|MN|}} \exp \left(-\frac{1}{2} \beta \sum_{(x,y) \in V(M,N)} (g_{x,y} - f_{x,y})^2 \right). \quad (4)$$

Here, we remark that α and β are referred to as hyperparameters in the statistical machine learning theory. α is an interaction between the nearest neighbour pairs of pixels. Equation (4) means that the additive white Gaussian noise is assumed as a degradation process, and β corresponds to the inverse of the variance of the additive White Gaussian noise.

The prior probability density function $\rho^{(M,N)}(f|\hat{\alpha}, \hat{\gamma})$ in Eq. (2) can be rewritten as

$$\rho^{(M,N)}(f|\alpha, \gamma) = \sqrt{\frac{\det(\gamma \mathbf{I}(M, N) + \alpha \mathbf{C}(M, N))}{(2\pi)^{MN}}} \exp \left(\frac{1}{2} f^T (\gamma \mathbf{I}(M, N) + \alpha \mathbf{C}(M, N)) f \right). \quad (5)$$

Here $\mathbf{C}(M, N)$ and $\mathbf{I}(M, N)$ denote an $MN \times MN$ matrix whose $(x, y|x', y')$ -components $\langle x, y | \mathbf{C}(M, N) | x', y' \rangle$ and $\langle x, y | \mathbf{I}(M, N) | x', y' \rangle$ ($(x, y) \in V(M, N)$, $(x', y') \in V(M, N)$) are defined by the following two-dimensional representations as

$$\langle x, y | \mathbf{C}(M, N) | x', y' \rangle \equiv 4\delta_{x',x}\delta_{y',y} - \delta_{x',x-1}\delta_{y',y} - \delta_{x',x+1}\delta_{y',y} - \delta_{x',x}\delta_{y',y-1} - \delta_{x',x}\delta_{y',y+1} \\ (x, x' = 0, 1, \dots, M-1; y, y' = 0, 1, \dots, N-1), \quad (6)$$

and

$$\langle x, y | \mathbf{I}(M, N) | x', y' \rangle \equiv \delta_{x', x} \delta_{y', y} (x, x' = 0, 1, \dots, M - 1; y, y' = 0, 1, \dots, N - 1), \quad (7)$$

respectively. Here, we remark that $\langle x, y | \mathbf{C}(M, N) | x', y' \rangle$ refers to the (i, i') -elements $(\mathbf{C}(M, N))_{i, i'}$ of the matrix $\mathbf{C}(M, N)$ for the relationships $r_i = (x, y)$ and $r_{i'} = (x', y')$ by Eq. (1). Moreover, $\langle x, y | \mathbf{I}(M, N) | x', y' \rangle$ denotes the (i, i') -elements $(\mathbf{I}(M, N))_{i, i'}$ of matrix $\mathbf{I}(M, N)$ in similar arguments. Hereafter, we use similar representations for the elements of the other matrix.

We introduce an $MN \times MN$ unitary matrix $\mathbf{U}(M, N)$ and its conjugate matrix $\mathbf{U}^\dagger(M, N)$ as follows:

$$\langle x, y | \mathbf{U}(M, N) | k, l \rangle \equiv \frac{1}{\sqrt{MN}} \exp \left(-i \frac{2\pi kx}{M} - i \frac{2\pi ly}{N} \right), \quad (8)$$

$$\langle k, l | \mathbf{U}^\dagger(M, N) | x, y \rangle \equiv \frac{1}{\sqrt{MN}} \exp \left(+i \frac{2\pi kx}{M} + i \frac{2\pi ly}{N} \right). \quad (9)$$

where $\mathbf{U}^\dagger(M, N) \equiv \overline{\mathbf{U}(M, N)}^T$ is the adjoint matrix of $\mathbf{U}(M, N)$. We can confirm the following relations for $\mathbf{U}(M, N)$ and $\mathbf{U}^\dagger(M, N)$:

$$\begin{aligned} \sum_{(k, l) \in W(M, N)} \langle x, y | \mathbf{U}(M, N) | k, l \rangle \langle k, l | \mathbf{U}^\dagger(M, N) | x', y' \rangle \\ = \delta_{x, x'} \delta_{y, y'} ((x, y), (x', y') \in V(M, N)), \end{aligned} \quad (10)$$

$$\begin{aligned} \sum_{(x, y) \in V(M, N)} \langle k, l | \mathbf{U}^\dagger(M, N) | x, y \rangle \langle x, y | \mathbf{U}(M, N) | k', l' \rangle \\ = \delta_{k, k'} \delta_{l, l'} ((k, l), (k', l') \in W(M, N)), \end{aligned} \quad (11)$$

where

$$\begin{aligned} W(M, N) \equiv \left\{ (k, l) \left| k = -\left\lfloor \frac{M-1}{2} \right\rfloor, -\left\lfloor \frac{M-1}{2} \right\rfloor + 1, \dots, +\left\lfloor \frac{M}{2} \right\rfloor, \right. \right. \\ \left. \left. l = -\left\lfloor \frac{N-1}{2} \right\rfloor, -\left\lfloor \frac{N-1}{2} \right\rfloor + 1, \dots, +\left\lfloor \frac{N}{2} \right\rfloor \right\}. \end{aligned} \quad (12)$$

We remark that $W(M, N)$ can be also regarded as an ordered set by considering the two-dimensional wave-number-dependent vector (k, l) as

$$p_i \equiv \left(-\left\lfloor \frac{M-1}{2} \right\rfloor + (i-1) \bmod(M), -\left\lfloor \frac{N-1}{2} \right\rfloor + \left\lfloor \frac{i-1}{M} \right\rfloor \right) \quad (i = 1, 2, 3, \dots, MN), \quad (13)$$

in the similar arguments for Eqs. (1). Equations (10) and (11) mean that $U(M, N)U^\dagger(M, N) = U^\dagger(M, N)U(M, N) = I(M, N)$, where $I(M, N)$ is an $MN \times MN$ identity matrix such that we have $U^\dagger(M, N) = U^{-1}(M, N)$. We confirm that $C(M, N)$ is diagonalized by the unitary matrix $U(M, N)$ as follows:

$$C(M, N) = U(M, N)\Lambda(M, N)U^\dagger(M, N), \quad (14)$$

where

$$\langle k, l | \Lambda(M, N) | k', l' \rangle \equiv \delta_{k, k'} \delta_{l, l'} \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right), \quad (15)$$

$$\lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \equiv 4 - 2\cos \left(\frac{2\pi k}{M} \right) - 2\cos \left(\frac{2\pi l}{N} \right). \quad (16)$$

The prior probability density function can be rewritten as

$$\begin{aligned} R^{(M, N)}(f | \alpha, \gamma) &= \sqrt{\frac{1}{(2\pi)^{MN}} \prod_{(k, l) \in W(M, N)} \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \\ &\times \exp \left(-\frac{1}{2} \sum_{(k, l) \in W(M, N)} f^T U^\dagger(M, N) | k, l \rangle \right. \\ &\times \left. \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \langle k, l | U(M, N) f \right). \end{aligned} \quad (17)$$

Here we remark that $U(M, N)f$ and $f^T U^\dagger(M, N)$ are the column vector and row vector, respectively. Also $\langle k, l | U(M, N)f$ and $f^T U^\dagger(M, N) | k, l \rangle$ are their i -th elements $(U(M, N)f)_i$ and $(f^T U^\dagger(M, N))_i$ with the relationship $p_i = (k, l)$ as per Eq. (13).

Now we introduce a new state vector F as

$$F \equiv \left(F \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) = \langle k, l | U(M, N) f \right)_{(k, l) \in W(M, N)}, \quad (18)$$

and a prior probability density function $\rho^{(M, N)}(F | \alpha, \gamma)$ for the state vector F :

$$\begin{aligned} R^{(M, N)}(F | \alpha, \gamma) &= \sqrt{\frac{1}{(2\pi)^{MN}} \prod_{(k, l) \in W(M, N)} \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \\ &\times \exp \left(-\frac{1}{2} \sum_{(k, l) \in W(M, N)} \left(\gamma \right. \right. \\ &\times \left. \left. + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) F \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) F \left(-\frac{2\pi k}{M}, -\frac{2\pi l}{N} \right) \right). \end{aligned} \quad (19)$$

By introducing a new state variable in a subspace as

$$\tilde{F} \left(\left(\frac{2\pi k}{M} \right) \left(\frac{M}{m} \right), \left(\frac{2\pi l}{N} \right) \left(\frac{N}{n} \right) \right) \equiv F \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)_{((k, l) \in W(m, n))}, \quad (20)$$

the marginal probability density function $R^{(m,n)}(\{\tilde{F}(\frac{2\pi k}{m}, \frac{2\pi l}{n})|(k, l) \in W(m, n)\}|\alpha, \gamma)$ in the momentum-space can be calculated as follows:

$$\begin{aligned} & R^{(m,n)}\left(\left\{\tilde{F}\left(\frac{2\pi k}{m}, \frac{2\pi l}{n}\right)\right|(k, l) \in W(m, n)\right)|\alpha, \gamma) \\ & \equiv \left(\prod_{(k,l) \in W(M,N)} \int dF\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right) \\ & \times \left(\prod_{(k,l) \in W(m,n)} \delta\left(F\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right) - \tilde{F}\left(\frac{2\pi k}{m}, \frac{2\pi l}{n}\right)\right)\right) R^{(M,N)}(F|\alpha, \gamma) \\ & = \sqrt{\frac{1}{(2\pi)^{mn}}} \prod_{(k,l) \in W(m,n)} \left(\gamma + \alpha \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right) \\ & \times \exp\left(-\frac{1}{2} \sum_{(k,l) \in W(m,n)} \left(\gamma + \alpha \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right) \tilde{F}\left(\frac{2\pi k}{m}, \frac{2\pi l}{n}\right) \tilde{F}\left(-\frac{2\pi k}{m}, -\frac{2\pi l}{n}\right)\right). \end{aligned} \quad (21)$$

We remark that the high-frequency modes are not ignored but are marginalized. Only low-frequency modes are rescaled or extended to recover the same frequency range as that of the original probabilistic Gaussian graphical model. The marginalization is referred to as the “trace Out” in the RG techniques in the physics. Again, we remark again that the present algorithm does the trace out; rather, it merely rescales instead of just ignoring the high-frequency modes. We consider that it is based on the standard RG.

Now we set new positive integers m and n with $m \leq M$ and $n \leq N$, respectively, and introduce the following scale transformation from the state vectors f in the space $V(M, N)$ to the new state vectors $\tilde{f} = (\tilde{f}_{x,y})|(x, y) \in V(m, n))$

$$\tilde{f}_{x,y} \equiv \sum_{(k,l) \in W(m,n)} \langle x, y | U(m, n) | k, l \rangle \tilde{F}\left(\left(\frac{2\pi k}{M}\right)\left(\frac{M}{m}\right), \left(\frac{2\pi l}{N}\right)\left(\frac{N}{n}\right)\right) ((x, y) \in V(m, n)), \quad (22)$$

such that

$$\tilde{f} \equiv U^\dagger(m, n) B(m, n | M, N) U(M, N) f, \quad (23)$$

where

$$\langle k', l' | B(m, n | M, N) | k, l \rangle \equiv \delta_{k,k'} \delta_{l,l'} \quad ((k', l') \in W(m, n), (k, l) \in W(M, N)), \quad (24)$$

$$\begin{aligned} \langle k, l | U(m, n) | x, y \rangle & \equiv \frac{1}{\sqrt{mn}} \exp\left(-i \frac{2\pi kx}{m} - i \frac{2\pi ly}{n}\right) \\ & ((x, y) \in V(m, n), (k, l) \in W(m, n)), \end{aligned} \quad (25)$$

Fig. 1 Visualization of the MSRG transformation of the prior probability density function of Eq. (26)

Fig. 2 Original images f^* ($M = N = 256$). a Parrots. b Couple. c Mandrill

and $U^\dagger(m, n) \equiv \overline{U(m, n)}^T$ is the adjoint matrix of $U(m, n)$. Equation (23) is one of MSRG transformations and is visualized as Fig. 1. Some examples obtained by applying the MSRG transformation in Eq. (23) to the original images f^* in Fig. 2 are shown in Figs. 3, 4 and 5.

We assume that the prior probability density function of $\tilde{f}$ is given as

Fig. 3 Reduced images $\tilde{f}^* \equiv U^\dagger(m, n)B(m, n|M, N)U(M, N)f^*$ for $m = n = 192$, such that $\frac{m}{M} = \frac{n}{N} = \frac{3}{4}$ for each original image f^* in Fig. 1. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 4 Reduced images $\tilde{f}^* \equiv U^\dagger(m, n)B(m, n|M, N)U(M, N)f^*$ for $m = n = 128$, such that $\frac{m}{M} = \frac{n}{N} = \frac{1}{2}$ for each original image f^* in Fig. 1. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 5 Reduced images $\tilde{f}^* \equiv U^\dagger(m, n)B(m, n|M, N)U(M, N)f^*$ for $m = n = 64$, such that $\frac{m}{M} = \frac{n}{N} = \frac{1}{4}$ for each original image f^* in Fig. 1. **a** Parrots. **b** Couple. **c** Mandrill

$$\begin{aligned} \rho^{(m, n)}(\tilde{f}|\alpha, \gamma) &= \sqrt{\frac{1}{(2\pi)^{mn}} \prod_{(k, l) \in W(m, n)} \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \\ &\times \exp \left(-\frac{1}{2} \sum_{(k, l) \in W(m, n)} \tilde{f}^T U^\dagger(m, n) |k, l \right) \\ &\times \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \langle k, l | U(m, n) \tilde{f} \rangle. \end{aligned} \quad (26)$$

The conditional probability density function $\rho^{(M,N)}(\mathbf{g}|\mathbf{f}, \beta)$ in Eq. (4) can be rewritten as

$$\rho^{(M,N)}(\mathbf{g}|\mathbf{f}, \beta) = \sqrt{\left(\frac{\beta}{2\pi}\right)^{MN}} \times \exp\left(-\frac{1}{2}\beta \sum_{(k,l) \in W(M,N)} (\mathbf{g} - \mathbf{f})^T \mathbf{U}^\dagger(M,N) |k, l\rangle \langle k, l| \mathbf{U}(M,N) (\mathbf{g} - \mathbf{f})\right). \quad (27)$$

Now we introduce a new state vector $\mathbf{F}$ as

$$\mathbf{G} \equiv \left(G\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right) = \langle k, l | \mathbf{U}(M,N) \mathbf{g} \right)_{(k,l) \in W(M,N)} \quad ((k, l) \in W(M, N)), \quad (28)$$

and the conditional probability density function $\rho^{(M,N)}(\mathbf{g}|\mathbf{f}, \beta)$ in Eq. (4) can be regarded as the conditional probability density function $R^{(M,N)}(\mathbf{G}|\mathbf{F}, \beta)$ in the momentum-space:

$$R^{(M,N)}(\mathbf{G}|\mathbf{F}, \beta) = \sqrt{\left(\frac{\beta}{2\pi}\right)^{MN}} \times \exp\left(-\frac{1}{2}\beta \sum_{(k,l) \in W(M,N)} \left(G\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right) - F\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right) \times \left(G\left(-\frac{2\pi k}{M}, -\frac{2\pi l}{N}\right) - F\left(-\frac{2\pi k}{M}, -\frac{2\pi l}{N}\right)\right)\right). \quad (29)$$

We introduce the scale transformation

$$\tilde{G}\left(\left(\frac{2\pi k}{M}\right)\left(\frac{M}{m}\right), \left(\frac{2\pi l}{N}\right)\left(\frac{N}{n}\right)\right) \equiv G\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right) \quad (30)$$

Fig. 6 Degraded images $\mathbf{g}^*$ generated from the original images $\mathbf{f}^*$ in Fig. 2 by according to the additive white Gaussian noise with average 0 and variance $1/\beta^* = 40^2$ ($M = N = 256$). **a** Parrots. **b** Couple. **c** Mandrill

Fig. 7 Reduced images $\tilde{g}^* \equiv U^\dagger(m, n)B(m, n|M, N)U(M, N)g^*$ for $m = n = 192$, such that $\frac{m}{M} = \frac{n}{N} = \frac{3}{4}$ for each degraded image g^* in Fig. 6. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 8 Reduced images $\tilde{g}^* \equiv U^\dagger(m, n)B(m, n|M, N)U(M, N)g^*$ for $m = n = 128$, such that $\frac{m}{M} = \frac{n}{N} = \frac{1}{2}$ for each degraded image g^* in Fig. 6. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 9 Reduced images $\tilde{g}^* \equiv U^\dagger(m, n)B(m, n|M, N)U(M, N)g^*$ for $m = n = 64$, such that $\frac{m}{M} = \frac{n}{N} = \frac{1}{4}$ for each degraded image g^* in Fig. 6. **a** Parrots. **b** Couple. **c** Mandrill

in the momentum-space, and the marginal probability density function $R^{(m, n)}((\tilde{G}(\frac{2\pi k}{m}, \frac{2\pi l}{n})|(k, l) \in W(m, n))|\{\tilde{F}(\frac{2\pi k}{m}, \frac{2\pi l}{n})|(k, l) \in W(m, n)\}, \beta)$ in the momentum space can be calculated as follows:

$$\begin{aligned}
& R^{(m,n)} \left(\left(\tilde{G} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) \right)_{(k,l) \in W(m,n)} \left| \left\{ \tilde{F} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) \right\}_{(k,l) \in W(m,n)} \right., \beta \right) \\
& \equiv \left(\prod_{(k,l) \in W(M,N) \setminus W(m,n)} \int dG \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \int dF \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
& \quad \times \left(\prod_{(k,l) \in W(m,n)} \delta \left(G \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) - \tilde{G} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) \right) \right) \\
& \quad \times \left(\prod_{(k,l) \in W(m,n)} \delta \left(F \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) - \tilde{F} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) \right) \right) R^{(M,N)}(G|F, \beta) \\
& = \sqrt{\left(\frac{\beta}{2\pi} \right)^{mn}} \exp \left(-\frac{1}{2} \beta \sum_{(k,l) \in W(m,n)} \left(\tilde{G} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) - \tilde{F} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) \right) \right. \\
& \quad \left. \times \left(\tilde{G} \left(-\frac{2\pi k}{m}, -\frac{2\pi l}{n} \right) - \tilde{F} \left(-\frac{2\pi k}{m}, -\frac{2\pi l}{n} \right) \right) \right). \tag{31}
\end{aligned}$$

We introduce the following scale transformation from the state vector $\mathbf{g}$ in the space $V(M, N)$ to new state vectors $\tilde{\mathbf{g}} = (\tilde{g}_{x,y})_{(x,y) \in V(m,n)}$

$$\tilde{g}_{x,y} \equiv \sum_{(k,l) \in W(m,n)} \langle x, y | U(m, n) | k, l \rangle \tilde{G} \left(\frac{2\pi k}{m}, \frac{2\pi l}{n} \right) \quad ((x, y) \in V(m, n)), \tag{32}$$

such that

$$\tilde{\mathbf{g}} \equiv U^\dagger(m, n) \mathbf{B}(m, n | M, N) U(M, N) \mathbf{g}. \tag{33}$$

Some examples obtained by applying the MSRG transformation in Eq. (33) to degraded images $\mathbf{g}$ in Fig. 6 are shown in Figs. 7, 8 and 9.

Moreover, we introduce the conditional probability density function of the degraded image $\tilde{\mathbf{g}} \equiv (\tilde{g}_{x,y})_{(x,y) \in V(m,n)}$, where the original image $\tilde{\mathbf{f}}$ given in the reduced space $V(m, n)$, is assumed to satisfy

$$\begin{aligned}
\rho^{(m,n)}(\tilde{\mathbf{g}} | \tilde{\mathbf{f}}, \beta) &= \sqrt{\left(\frac{\beta}{2\pi} \right)^{mn}} \\
& \quad \times \exp \left(-\frac{1}{2} \beta \sum_{(k,l) \in W(m,n)} \left(\tilde{\mathbf{g}} - \tilde{\mathbf{f}} \right)^\top U^\dagger(m, n) | k, l \rangle \langle k, l | U(m, n) \left(\tilde{\mathbf{g}} - \tilde{\mathbf{f}} \right) \right). \tag{34}
\end{aligned}$$

3 Momentum-Space Renormalization Group Approaches in EM Algorithm

In this section, we provide a practical scheme of the EM algorithm in the MSRG approaches. The fundamental framework is based on the maximization of marginal likelihood. We first formulate the probability density function of $\tilde{\mathbf{g}}$ and regard it as a marginal likelihood function of α , β and γ .

Under some assumptions in the previous section, the marginal likelihood $\rho^{(m,n)}(\tilde{\mathbf{g}}|\alpha, \beta, \gamma)$ in the space $V(m, n)$ is defined by

$$\rho^{(m,n)}(\tilde{\mathbf{g}}|\alpha, \beta, \gamma) \equiv \int \rho^{(m,n)}(\tilde{\mathbf{g}}|\tilde{\mathbf{f}}, \beta) \rho^{(m,n)}(\tilde{\mathbf{f}}|\alpha, \gamma) d\tilde{\mathbf{f}}. \quad (35)$$

Our proposed framework is designed to achieve the estimation of the hyperparameters α , β and γ by maximizing $\rho^{(m,n)}(\tilde{\mathbf{g}}|\alpha, \beta, \gamma)$ as follows:

$$\begin{aligned} \left(\frac{1}{mn}\right) \ln(\rho^{(m,n)}(\tilde{\mathbf{g}}|\alpha, \beta, \gamma)) &= -\frac{1}{2} \ln\left(\frac{2\pi}{\beta}\right) \\ &+ \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln\left(\gamma + \alpha \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right) \\ &- \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln\left(\beta + \gamma + \alpha \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right) \\ &- \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \tilde{\mathbf{g}}^T \mathbf{U}^\dagger(m, n)|k, l \rangle \frac{\beta\left(\gamma + \alpha \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)}{\left(\beta + \gamma + \alpha \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)} \langle k, l | \mathbf{U}(m, n) \tilde{\mathbf{g}}. \end{aligned} \quad (36)$$

Estimates of hyperparameters, $\hat{\alpha}_{m,n}(\mathbf{g})$, $\hat{\beta}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$, are determined so as to maximize the marginal likelihood $\rho^{(m,n)}(\tilde{\mathbf{g}} = \mathbf{U}^\dagger(m, n)\mathbf{B}(m, n|M, N)\mathbf{U}(M, N)\mathbf{g}|\alpha, \beta, \gamma)$ with respect to α , β and γ , respectively, as follows:

$$\begin{aligned} &\left(\hat{\alpha}_{m,n}(\mathbf{g}), \hat{\beta}_{m,n}(\mathbf{g}), \hat{\gamma}_{m,n}(\mathbf{g})\right) \\ &= \arg \max_{(\alpha, \beta, \gamma)} \rho^{(m,n)}\left(\tilde{\mathbf{g}} = \mathbf{U}^\dagger(m, n)\mathbf{B}(m, n|M, N)\mathbf{U}(M, N)\mathbf{g}|\alpha, \beta, \gamma\right). \end{aligned} \quad (37)$$

By considering the extremum conditions of $\rho^{(m,n)}(\tilde{\mathbf{g}} = \mathbf{U}^\dagger(m, n)\mathbf{B}(m, n|M, N)\mathbf{U}(M, N)\mathbf{g}|\alpha, \beta, \gamma)$ with respect to α , β and γ , the simultaneous deterministic equations for the estimates of hyperparameters, $\hat{\alpha}_{m,n}(\mathbf{g})$, $\hat{\beta}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$, are derived as follows:

$$\begin{aligned}
& \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{\lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)}{\hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)} \\
&= \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{\lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)}{\hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)} \\
&+ \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \tilde{\mathbf{g}}^T \mathbf{U}^\dagger(m, n) |k, l\rangle \\
&\times \frac{\hat{\beta}_{m,n}(\mathbf{g})^2 \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)}{\left(\hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)^2} \langle k, l | \mathbf{U}(m, n) \tilde{\mathbf{g}}, \quad (38)
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{1}{\hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)} \\
&= \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{1}{\hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)} \\
&+ \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \tilde{\mathbf{g}}^T \mathbf{U}^\dagger(m, n) |k, l\rangle \\
&\times \frac{\hat{\beta}_{m,n}(\mathbf{g})^2}{\left(\hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)^2} \langle k, l | \mathbf{U}(m, n) \tilde{\mathbf{g}}, \quad (39)
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\hat{\beta}_{m,n}(\mathbf{g})} &= \frac{1}{mn} \sum_{(k,l) \in W(m,n)} \frac{1}{\hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)} \\
&+ \frac{1}{mn} \sum_{(k,l) \in W(m,n)} \tilde{\mathbf{g}}^T \mathbf{U}^\dagger(m, n) |k, l\rangle \\
&\times \frac{\left(\hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)^2}{\left(\hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g}) + \hat{\alpha}_{m,n}(\mathbf{g}) \lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)^2} \langle k, l | \mathbf{U}(m, n) \tilde{\mathbf{g}}. \quad (40)
\end{aligned}$$

The practical algorithm to solve Eqs. (38), (39) and (40) with respect to $\hat{\alpha}_{m,n}(\mathbf{g})$, $\hat{\beta}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$ is given as follows:

4 Hyperparameter Estimation and Noise Reduction Algorithm

Step 1: Input a given data point $\mathbf{g}$ and the set of positive integers (m, n) . Compute $\tilde{G}(k, l)$ using

$$\tilde{G}(k, l) \leftarrow \frac{1}{\sqrt{mn}} \sum_{(x,y) \in V(m,n)} \mathbf{g}^T \mathbf{B}^T(m, n | M, N) |x, y \rangle \exp \left(-i \frac{2\pi kx}{m} - i \frac{2\pi ly}{n} \right) \quad (41)$$

$$((k, l) \in W(m, n)).$$

Set $\alpha(t)$, $\beta(t)$ and $\gamma(t)$ as initial values, and $t \leftarrow 0$.

Step 2: Update t by $t \leftarrow t + 1$ and compute u , v and $\beta(t)$ from $\alpha(t - 1)$, $\beta(t - 1)$ and $\gamma(t - 1)$ as follows:

$$u \leftarrow \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{\lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)}{\left(\beta(t - 1) + \gamma(t - 1) + \alpha(t - 1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}$$

$$+ \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} |\tilde{G}(k, l)|^2 \frac{\beta(t - 1)^2 \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)}{\left(\left(\beta(t - 1) + \gamma(t - 1) + \alpha(t - 1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \right)^2}, \quad (42)$$

Fig. 10 Restored images $\hat{f}(M, N)$ for the degraded images $\mathbf{g}^*$ in Fig. 6, such that $m = M$ and $n = N$. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 11 Restored images $\hat{f}(m, n)$ for the degraded images g^* in Fig. 6 in the case of $m = n = 256$ such that $\frac{m}{M} = \frac{n}{N} = \frac{3}{4}$. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 12 Restored images $\hat{f}(m, n)$ for the degraded images g^* in Fig. 6 in the case of $m = n = 256$ such that $\frac{m}{M} = \frac{n}{N} = \frac{1}{2}$. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 13 Restored images $\hat{f}(m, n)$ for the degraded images g^* in Fig. 6 in the case of $m = n = 64$ such that $\frac{m}{M} = \frac{n}{N} = \frac{1}{4}$. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 14 Computational times for the estimation processes of hyperparameters, $\hat{\alpha}_{n,n}(\mathbf{g})$, $\hat{\beta}_{n,n}(\mathbf{g})$ and $\hat{\gamma}_{n,n}(\mathbf{g})$, in the numerical experiments for the degraded images $\mathbf{g}^*$ in Fig. 6. **a** Parrots. **b** Couple. **c** Mandrill

$$v \Leftarrow \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{1}{\left(\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} + \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} |\tilde{G}(k,l)|^2 \frac{\beta(t-1)^2}{\left(\left(\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \right)^2}, \quad (43)$$

$$\beta(t) \Leftarrow \left(\frac{1}{mn} \sum_{(k,l) \in W(m,n)} \frac{1}{\left(\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} + \frac{1}{mn} \sum_{(k,l) \in W(m,n)} |\tilde{G}(k,l)|^2 \frac{\left(\gamma(t-1) + \alpha(t-1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)^2}{\left(\beta(t-1) + \gamma(t-1) + \alpha(t-1) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)^2} \right)^{-1}. \quad (44)$$

Fig. 15 Logarithm of the signal-to-noise ratio $10\log_{10}(\frac{\text{Var}[f^*]}{\|\hat{f}(n,n) - f^*\|^2})(\text{dB})$. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

Step 3: Repeat the following updates for $\alpha(t)$ and $\gamma(t)$ until they converge:

$$\begin{pmatrix} \alpha(t) \\ \gamma(t) \end{pmatrix} \Leftarrow \begin{pmatrix} \alpha(t) \left(\frac{1}{2mnu} \sum_{(k,l) \in W(m,n)} \frac{\lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)}{\left(\gamma(t) + \alpha(t)\lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)} \right)^{1/2} \\ \gamma(t) \left(\frac{1}{2mnv} \sum_{(k,l) \in W(m,n)} \frac{1}{\left(\gamma(t) + \alpha(t)\lambda\left(\frac{2\pi k}{M}, \frac{2\pi l}{N}\right)\right)} \right)^{1/2} \end{pmatrix}. \quad (45)$$

Step 4: Set $\hat{\alpha}_{m,n}(\mathbf{g}) \Leftarrow \alpha(t)$, $\hat{\beta}_{m,n}(\mathbf{g}) \Leftarrow \beta(t)$, $\hat{\gamma}_{m,n}(\mathbf{g}) \Leftarrow \gamma(t)$. Go to **Step 5** if $\hat{\alpha}_{m,n}(\mathbf{g})$, $\hat{\beta}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$ converge, and return to **Step 2** otherwise.

Step 5: Repeat the following update rule until $\hat{f}(\hat{\alpha}_{m,n}(\mathbf{g}), \hat{\beta}_{m,n}(\mathbf{g}), \hat{\gamma}(\mathbf{g}))$ converges:

$$\begin{aligned} \hat{f}(\hat{\alpha}_{m,n}(\mathbf{g}), \hat{\beta}_{m,n}(\mathbf{g}), \hat{\gamma}(\mathbf{g})) &\Leftarrow \frac{1}{4\hat{\alpha}_{m,n}(\mathbf{g}) + \hat{\beta}_{m,n}(\mathbf{g}) + \hat{\gamma}_{m,n}(\mathbf{g})} \\ &\times \left(\hat{\beta}_{m,n}(\mathbf{g})\mathbf{g} + (\hat{\gamma}_{m,n}(\mathbf{g})\mathbf{I}(M, N) \right. \\ &\left. + \hat{\alpha}_{m,n}(\mathbf{g})(4\mathbf{I}(M, N) - \mathbf{C}(M, N)))\hat{f}(\hat{\alpha}_{m,n}(\mathbf{g}), \hat{\beta}_{m,n}(\mathbf{g}), \hat{\gamma}) \right). \end{aligned} \quad (46)$$

Some restored results $\hat{f}(\hat{\alpha}_{m,n}(\mathbf{g}^*), \hat{\beta}_{m,n}(\mathbf{g}^*), \hat{\gamma}(\mathbf{g}^*))$ obtained by applying the above algorithm to degraded images $\mathbf{g}^*$ in Fig. 6 are shown in Figs. 10, 11, 12 and 13. Computational times in the estimation processes of hyperparameters, $\hat{\alpha}_{m,n}(\mathbf{g})$, $\hat{\beta}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$, in the numerical experiments for the degraded images $\mathbf{g}^*$ in Fig. 6 are also shown in Fig. 14. Numerical experimental results of the logarithm of the signal-to-noise ratio $10\log_{10}\left(\frac{\text{Var}[f^*]}{\|\hat{f}(n,n) - f^*\|^2}\right)$ (dB) for the degraded images $\mathbf{g}^*$ in Fig. 6 are shown in Fig. 15. Here, $\text{Var}[f^*]$ is a variance of the light intensities for each colour, red, green and blue, on all the pixels in the original image f^* .

5 Statistical Performance Estimation of Gaussian Graphical Model

In this section, we estimate the statistical performance of our framework of the maximization of the renormalized marginal likelihood $\rho^{(m,n)}(\tilde{\mathbf{g}}|\alpha, \beta, \gamma)$ in Eq. (36) with Eq. (33). In the present Bayesian inference method, we assume that the data vectors $\mathbf{g}$ are generated from the conditional probability density function $\rho^{(M,N)}(\mathbf{g}|\mathbf{f}, \beta^*)$ in Eq. (4), under the assumption that a parameter vector $\mathbf{f} = \mathbf{f}^*$ is given.

$$\begin{cases} \bar{\alpha}_{m,n}(\mathbf{f}^*) \equiv \int \hat{\alpha}_{m,n}(\mathbf{g}) \rho^{(M,N)}(\mathbf{g}|\mathbf{f} = \mathbf{f}^*, \beta^*) d\mathbf{g}, \\ \bar{\beta}_{m,n}(\mathbf{f}^*) \equiv \int \hat{\beta}_{m,n}(\mathbf{g}) \rho^{(M,N)}(\mathbf{g}|\mathbf{f} = \mathbf{f}^*, \beta^*) d\mathbf{g}, \\ \bar{\gamma}_{m,n}(\mathbf{f}^*) \equiv \int \hat{\gamma}_{m,n}(\mathbf{g}) \rho^{(M,N)}(\mathbf{g}|\mathbf{f} = \mathbf{f}^*, \beta^*) d\mathbf{g}. \end{cases} \quad (47)$$

Equation (47) provides the statistical average of the hyperparameters in the EM algorithm with respect to the probability density function $\rho^{(M,N)}(\mathbf{g}|\mathbf{f} = \mathbf{f}^*, \beta^*)$. For example, suppose that we generate D degraded images $\mathbf{g}^{(1)}(\mathbf{f}^*)$, $\mathbf{g}^{(2)}(\mathbf{f}^*)$, ..., $\mathbf{g}^{(D)}(\mathbf{f}^*)$ from the given original image $\mathbf{f}^*$, and calculate the estimates of hyperparameters for each degraded image $\mathbf{g}^{(d)}(\mathbf{f}^*)$ ($d = 1, 2, \dots, D$). Then we can estimate the statistical average of the estimates in the EM algorithm by the sample average of the estimates obtained by applying the EM algorithm to each degraded image $\mathbf{g}^{(d)}(\mathbf{f}^*)$. Equation (47) can be regarded as the statistical average of estimates in the EM algorithm with respect to the infinite number of degraded images generated from the probability density function $\rho^{(M,N)}(\mathbf{g}|\mathbf{f} = \mathbf{f}^*, \beta^*)$ as follows:

$$\begin{cases} \bar{\alpha}_{m,n}(\mathbf{f}^*) \simeq \frac{1}{D} \sum_{d=1}^D \hat{\alpha}_{m,n}(\mathbf{g}^{(d)}(\mathbf{f}^*)), \\ \bar{\beta}_{m,n}(\mathbf{f}^*) \simeq \frac{1}{D} \sum_{d=1}^D \hat{\beta}_{m,n}(\mathbf{g}^{(d)}(\mathbf{f}^*)), \\ \bar{\gamma}_{m,n}(\mathbf{f}^*) \simeq \frac{1}{D} \sum_{d=1}^D \hat{\gamma}_{m,n}(\mathbf{g}^{(d)}(\mathbf{f}^*)). \end{cases} \quad (48)$$

The detailed analysis for the EM algorithm in the case of $M = m$ and $N = n$ appears in Ref. [20].

We consider the average of the logarithm of the renormalized marginal likelihood $\ln(\rho^{(m,n)}(\tilde{\mathbf{g}} = \mathbf{U}^\dagger(m, n)\mathbf{B}(m, n|M, N)\mathbf{U}(M, N)\mathbf{g}|\alpha, \beta, \gamma))$ in the probability density function $\rho^{(M,N)}(\mathbf{g}|\mathbf{f}^*, \beta^*)$ as follows:

$$\begin{aligned} L_{m,n}(\alpha, \beta, \gamma|\beta^*, \mathbf{f}^*) &\equiv \frac{1}{mn} \int \rho^{(M,N)}(\mathbf{g}|\mathbf{f} = \mathbf{f}^*, \beta^*) \\ &\times \ln\left(\rho^{(m,n)}\left(\tilde{\mathbf{g}} = \mathbf{U}^\dagger(m, n)\mathbf{B}(m, n|M, N)\mathbf{U}(M, N)\mathbf{g}|\alpha, \beta, \gamma\right)\right) d\mathbf{g} \end{aligned} \quad (49)$$

By substituting Eq. (36) to Eq. (49), we can obtain the expression of $L_{m,n}(\alpha, \beta, \gamma|\beta^*, \mathbf{f}^*)$ in terms of the original image $\mathbf{f}^*$ as follows:

$$\begin{aligned}
L_{m,n}(\alpha, \beta, \gamma | \beta^*, f^*) &= \int \rho^{(M,N)}(\mathbf{g} | f = f^*, \beta^*) \left(-\frac{1}{2} \ln \left(\frac{2\pi}{\beta} \right) \right. \\
&\quad + \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
&\quad - \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
&\quad - \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \mathbf{g}^T \mathbf{U}^\dagger(M, N) \mathbf{B}^T(m, n | M, N) | k, l \rangle \\
&\quad \times \frac{\beta \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}{\left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \langle k, l | \mathbf{B}(m, n | M, N) \mathbf{U}(M, N) \mathbf{g} \rangle \Bigg) \\
&= -\frac{1}{2} \ln \left(\frac{2\pi}{\beta} \right) \\
&\quad + \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
&\quad - \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
&\quad - \frac{1}{2mn} \int \rho^{(M,N)}(\mathbf{g} | f = f^*, \beta^*) \\
&\quad \times \left(\sum_{(k,l) \in W(m,n)} \mathbf{g}^T \mathbf{U}^\dagger(M, N) \mathbf{B}^T(m, n | M, N) | k, l \rangle \right. \\
&\quad \times \frac{\beta \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}{\left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \langle k, l | \mathbf{B}(m, n | M, N) \mathbf{U}(M, N) \mathbf{g} \rangle \Bigg) \\
&= -\frac{1}{2} \ln \left(\frac{2\pi}{\beta} \right) + \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
&\quad - \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right) \\
&\quad - \frac{1}{2mn} \int \rho^{(M,N)}(\mathbf{g} | f = f^*, \beta^*) \\
&\quad \times \left(\sum_{(k,l) \in W(m,n)} (\mathbf{g} - f^*)^T \mathbf{U}^\dagger(M, N) \mathbf{B}^T(m, n | M, N) | k, l \rangle \right. \\
&\quad \times \frac{\beta \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}{\left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \langle k, l | \mathbf{B}(m, n | M, N) \mathbf{U}(M, N) (\mathbf{g} - f^*) \rangle \Bigg) d\mathbf{g} \\
&\quad - \frac{1}{2mn} \int \rho^{(M,N)}(\mathbf{g} | f = f^*, \beta^*) d\mathbf{g} \\
&\quad \times \left(\sum_{(k,l) \in W(m,n)} (f^*)^T \mathbf{U}^\dagger(M, N) \mathbf{B}^T(m, n | M, N) | k, l \rangle \right. \\
&\quad \times \frac{\beta \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}{\left(\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)} \langle k, l | \mathbf{B}(m, n | M, N) \mathbf{U}(M, N) f^* \rangle \Bigg). \tag{50}
\end{aligned}$$

By integrating the fourth and fifth term of Eq. (50) with respect to $\mathbf{g}$, we obtain

$$\begin{aligned} L_{m,n}(\alpha, \beta, \gamma | \beta^*, \mathbf{f}^*) &\equiv \frac{1}{mn} \int \rho^{(M,N)}(\mathbf{g} | \mathbf{f} = \mathbf{f}^*, \beta^*) \\ &\quad \times \ln \left(\rho^{(m,n)} \left(\tilde{\mathbf{g}} = \mathbf{U}^\dagger(m, n) \mathbf{B}(m, n | M, N) \mathbf{U}(M, N) \mathbf{g} \mid \alpha, \beta, \gamma \right) \right) d\mathbf{g} \\ &= -\frac{1}{2} \ln \left(\frac{2\pi}{\beta} \right) + \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \ln \left(\frac{\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)}{\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)} \right) \\ &\quad - \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} \frac{\beta \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}{\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)} \\ &\quad - \frac{1}{2mn} \sum_{(k,l) \in W(m,n)} (\mathbf{f}^*)^\top \mathbf{U}^\dagger(N) \mathbf{B}^\top(m, n | M, N) |k, l\rangle \\ &\quad \times \frac{\beta \left(\gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right) \right)}{\beta + \gamma + \alpha \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)} \langle k, l | \mathbf{B}(m, n | M, N) \mathbf{U}(M, N) \mathbf{f}^*. \end{aligned} \quad (51)$$

The statistical averages for the estimates of the hyperparameters are given as

$$\left(\bar{\alpha}_{m,n}(\mathbf{f}^*), \bar{\beta}_{m,n}(\mathbf{f}^*), \bar{\gamma}_{m,n}(\mathbf{f}^*) \right) = \arg \max_{(\alpha, \beta, \gamma)} L_{m,n}(\alpha, \beta, \gamma | \beta^*, \mathbf{f}^*). \quad (52)$$

The statistical performance of our proposed scheme is given by

$$\begin{aligned} D_{m,n}(\beta^*, \mathbf{f}^*) &\equiv \frac{1}{N^2} \int \left\| \mathbf{f} - \hat{\mathbf{f}} \left(\bar{\alpha}_{m,n}(\mathbf{f}^*), \bar{\beta}_{m,n}(\mathbf{f}^*), \bar{\gamma}_{m,n}(\mathbf{f}^*) \mid \mathbf{g} \right) \right\|^2 \\ &\quad \times \rho^{(M,N)}(\mathbf{g} | \mathbf{f} = \mathbf{f}^*, \beta^*) d\mathbf{g} \\ &= \frac{1}{MN} \sum_{(k,l) \in W(M,N)} \frac{1}{\beta^*} \left(\frac{\bar{\beta}_{m,n}(\mathbf{f}^*)}{\bar{\beta}_{m,n}(\mathbf{f}^*) + \bar{\gamma}_{m,n}(\mathbf{f}^*) + \bar{\alpha}_{m,n}(\mathbf{f}^*) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)} \right)^2 \\ &\quad + \frac{1}{MN} \sum_{(k,l) \in W(M,N)} (\mathbf{f}^*)^\top \mathbf{U}(M, N) |k, l\rangle \\ &\quad \times \left(\frac{\bar{\gamma}_{m,n}(\mathbf{f}^*) + \bar{\alpha}_{m,n}(\mathbf{f}^*) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)}{\bar{\beta}_{m,n}(\mathbf{f}^*) + \bar{\gamma}_{m,n}(\mathbf{f}^*) + \bar{\alpha}_{m,n}(\mathbf{f}^*) \lambda \left(\frac{2\pi k}{M}, \frac{2\pi l}{N} \right)} \right)^2 \langle k, l | \mathbf{U}(M, N) \mathbf{f}^*, \end{aligned} \quad (53)$$

where

$$\hat{f}_{x,y}(\alpha, \beta, \gamma | \mathbf{g}) = \int f_{x,y} \rho^{(M,N)}(\mathbf{f} | \mathbf{g}, \alpha, \beta, \gamma) d\mathbf{f}, \quad (54)$$

$$\rho^{(M,N)}(f|g, \alpha, \beta, \gamma) \equiv \frac{\rho^{(M,N)}(g|f, \beta) \rho^{(M,N)}(f|\alpha, \gamma)}{\rho^{(M,N)}(g|\alpha, \beta, \gamma)}. \quad (55)$$

Statistical averages of logarithm of the signal-to-noise ratio, $10\log_{10}(\frac{\text{Var}[f^*]}{D_{n,n}(\beta^*, f^*)})$ (dB) ($m = n$) in Eq. (53) for the original images f^* in Fig. 3 are shown in Fig. 16. Statistical averages of the hyperparameters, $\bar{\alpha}_{n,n}(f^*)$, $\bar{\beta}_{n,n}(f^*)$ and $\bar{\gamma}_{n,n}(f^*)$, are shown in Figs. 17, 18 and 19, respectively. Flows of estimates of the hyperparameters, $(\bar{\alpha}_{n,n}(f^*), \bar{\beta}_{n,n}(f^*))$ and $(\bar{\alpha}_{n,n}(f^*), \bar{\gamma}_{n,n}(f^*))$, in the momentum-space renormalization transformation in the EM algorithm of the Gaussian graphical model for the original images f^* in Fig. 3 are shown in Figs. 20 and 21. These flows in Fig. 20 are drawn

Fig. 16 Statistical average of the logarithm of the signal-to-noise ratio, $10\log_{10}(\frac{\text{Var}[f^*]}{D_{n,n}(\beta^*, f^*)})$ (dB) ($m = n$) in Eq. (53) for the original images in Fig. 3. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 17 Statistical average $\bar{\alpha}_{n,n} = \bar{\alpha}_{n,n}(f^*)$ in Eq. (52) for the original images in Fig. 3. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 18 Statistical average $\bar{\beta}_{n,n} = \bar{\beta}_{n,n}(f^*)$ in Eq. (52) for the original images in Fig. 3. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 19 Statistical average $\bar{\gamma}_{n,n} = \bar{\gamma}_{n,n}(f^*)$ in Eq. (52) for the original images in Fig. 3. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 20 Flows of $(\bar{\alpha}_{n,n}(f^*), \bar{\beta}_{n,n}(f^*))$ in the MSRG transformations in the EM algorithm of the Gaussian graphical model for the original images in Fig. 3. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

Fig. 21 Flows of $(\bar{\alpha}_{n,n}(f^*), \bar{\gamma}_{n,n}(f^*))$ in the MSRG transformations in the EM algorithm of the Gaussian graphical model for the original images in Fig. 3. The red, green and blue circles correspond to the results for each light colour intensity. **a** Parrots. **b** Couple. **c** Mandrill

by plotting $(\bar{\alpha}_{n,n}(f^*), \bar{\beta}_{n,n}(f^*))$ in Figs. 17 and 18 and the flows in Fig. 21 are drawn by plotting $(\bar{\alpha}_{n,n}(f^*), \bar{\gamma}_{n,n}(f^*))$ from Figs. 17 and 19.

For the limits of $\frac{n}{N} \rightarrow +0$ and $\frac{m}{M} \rightarrow +0$, $\rho^{(m,n)}(\tilde{f}|\alpha, \gamma)$ in Eq. (26) and $\rho^{(m,n)}(\tilde{g}|\tilde{f}, \alpha, \beta)$ in Eq. (34), are reduced to $\rho^{(m,n)}(\tilde{f}|\alpha = 0, \gamma)$ and $\rho^{(m,n)}(\tilde{g}|\tilde{f}, \beta)$, respectively. This means that the joint probability density function $\rho^{(m,n)}(\tilde{f}, \tilde{g}|\alpha, \beta, \gamma)$ has a fixed point at $(0, \beta, \gamma)$ in the scale transformations in Eqs. (23) and (33). In $\frac{n}{N} > \frac{1}{4}$ and $\frac{m}{M} > \frac{1}{4}$, the estimates, $\hat{\alpha}_{m,n}(\mathbf{g})$, $\hat{\beta}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$ are located along the flow by MSRG

transformation. In $\frac{n}{N} < \frac{1}{4}$ and $\frac{m}{M} < \frac{1}{4}$, they leave from the flow. In Fig. 9, it is seen that the noise in the reduced image $\tilde{\mathbf{g}}$ is hardly recognized for $\frac{n}{N} = \frac{1}{4}$ and $\frac{m}{M} = \frac{1}{4}$. In fact, the hyperparameter β corresponds to the inverse of the variance in the additive white Gaussian noise, and then $\hat{\beta}_{m,n}(\mathbf{g})$ rapidly increases although $\hat{\alpha}_{m,n}(\mathbf{g})$ and $\hat{\gamma}_{m,n}(\mathbf{g})$ remain near the renormalization flow within the limit of $\frac{n}{N} \rightarrow +0$ and $\frac{m}{M} \rightarrow +0$.

6 Concluding Remarks

In the present review paper, we have summarized the MSRG approaches for the Gaussian graphical models in a finite-size square grid graph. We have provided the formulations and a practical algorithm for noise reduction applications in image processing using the Bayesian sublinear computational time modeling. Moreover, we derive also the statistical performance schemes of the above systems. from the statistical–mechanical point of view. The estimated results of the hyperparameters are first located along the flow of the MSRG transformations in the region $\frac{mn}{MN} > (\frac{1}{4})^2$. However, the sublinear modeling cannot recognize the noise in the reduced image $\tilde{\mathbf{g}}$ in the region $\frac{mn}{MN} < (\frac{1}{4})^2$ and then the estimate of the hyperparameter β , which corresponds to the inverse of the variance in the additive white Gaussian noise, goes to infinity with $\frac{mn}{MN}$ decreasing in the region.

We consider finite-size system only, and our formulation is within the discrete Fourier transformation. However, if we consider a system size with infinity as the limit, such that $M \rightarrow +\infty$ and $N \rightarrow +\infty$, the state vector $\mathbf{f}$ would be composed of an infinite number of components. At this stage, $f_{x,y}$ is a function of x and y which are any real number in $(-\infty + \infty)$, and which can then be rewritten as $f(x, y)$. $\rho(\mathbf{f}|\alpha, \gamma)$ is a functional of $\mathbf{f} = \{f(x, y) | (x, y) \in (-\infty, +\infty)^2\}$ and is given by

$$\rho(\mathbf{f}|\alpha, \gamma) \propto \exp \left(-\frac{1}{2} \alpha \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\left(\frac{\partial}{\partial x} f(x, y) \right)^2 + \left(\frac{\partial}{\partial y} f(x, y) \right)^2 \right) dx dy - \frac{1}{2} \gamma \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)^2 dx dy \right), \quad (56)$$

instead of Eq. (2). For the probability density function in Eq. (56), the integral $\int(\cdots)d\mathbf{f}$ means a functional integral. This is a starting point of the statistical field theory in the physics [25–27]. Big data sciences are characterized by extremely large dimensional state vectors, and we expect that the statistical field theory should provide powerful applications in such cases. Indeed, this theory is one of the most exciting research paradigms for sublinear computational time modeling in the data sciences.

Acknowledgements This work was partly supported by the JST-CREST (No.JPMJCR1402) for Japan Science and Technology Agency and the JSPS KAKENHI Grant (nos. 18H03303 and 18K11459) for Ministry of Education, Culture, Sports, Science and Technology.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. New York: MIT Press.
2. Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. New York: MIT Press.
3. Oppor, M., & Saad, D. (Eds.). (2001). *Advanced Mean Field Methods – Theory and Practice* –. MIT Press.
4. Tanaka, K. (2002). Statistical–mechanical approach to image processing. *Journal of Physics A Mathematical and General*, 35(37), R81–R150.
5. Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7), 2282–2312.
6. Pelizzola, A. (2005). Cluster variation method in statistical physics and probabilistic graphical models. *Journal of Physics A Mathematical and General*, 38(33), R309–R339. (Topical Review).
7. Wainwright, M. J., & Jordan, M. I. (2008). *Graphical Models, Exponential Families, and Variational Inference*. : now Publishing Inc.
8. Mézard, M., & Montanari, A. (2009). *Information*. New York, USA: Physics and Computation Oxford University Press.
9. Wilson, K. G., & Kogut, J. (1974). The renormalization group and the ϵ expansion. *Physics Report (Section C of Physics Letters)*, 12(2), 75–200.
10. Fisher, M. E. (1974). The renormalization group in the theory of critical behavior. *Review of Modern Physics*, 46(4), 597–616.
11. Wilson, K. G. (1983). The renormalization group and critical phenomena. *Review of Modern Physics*, 55(9), 583–600.
12. Amit, D. J., & Martin-Mayor, V. (2005). *Field theory, the renormalization group, and critical phenomena: Graphs to computers*. Singapore: World Scientific.
13. Nishimori, H., & Ortiz, G. (2011). *Elements of phase transitions and critical phenomena*. Oxford: Oxford University Press.
14. Tanaka, K., Kataoka, S., Yasuda, M., & Ohzeki, M. (2015). Inverse renormalization group transformation in Bayesian image segmentations. *Journal of the Physical Society of Japan*, 84(4), Article ID: 045001.
15. Tanaka, K., Nakamura, M., Kataoka, S., Ohzeki, M., & Yasuda, M. (2018). Momentum-space renormalization group transformation in Bayesian image modeling by Gaussian graphical model. *Journal of the Physical Society of Japan*, 87(8), Article ID: 085001.
16. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, 39(1), 1–38. (with discussions).
17. Mengersen, K. L., Robert, C. P., & Titterton, D. M. (2011). *Mixtures: Estimation and applications*. Oxford: Wiley.
18. Geman, D. (1990). *Random fields and inverse problems in imaging* (Vol. 1427, pp. 113–193)., Lecture notes in mathematics Berlin: Springer.
19. Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Boca Raton: CRC.
20. Tanaka, K., & Titterton, D. M. (2007). Statistical trajectory of approximate EM algorithm for probabilistic image processing. *Journal of Physics A Mathematical and Theoretical*, 40(37), 11285–11300.
21. Kataoka, S., Yasuda, M., & Tanaka, K. (2012). Statistical analysis of Gaussian image inpainting problems. *Journal of the Physical Society of Japan*, 81(2), Article ID.025001.
22. Kataoka, S., Yasuda, M., Furtlehner, C., & Tanaka, K. (2014). Traffic data reconstruction based on Markov random field modeling. *Inverse Problems*, 30(2), Article ID: 025003.

23. Kuwatani, T., Nagata, K., Okada, M., & Toriumi, M. (2014). Markov-random-field modeling for linear seismic tomography. *Physical Review E*, 90(4), Article ID: 042137.
24. Kuwatani, T., Nagata, K., Okada, M., & Toriumi, M. (2014). Markov random field modeling for mapping geofluid distributions from seismic velocity structures. *Earth Planets and Space*, 66, Article ID: 5.
25. Parisi, G. (1988). *Statistical field theory*. Boston: Addison-Wesley.
26. Itzykson, C., & Drouffee, J.-M. (1989). *Statistical field theory: Volume 1 from brown motion to renormalization and lattice gauge theory*. Cambridge: Cambridge University Press.
27. Itzykson, C., & Drouffee, J.-M. (1989). *Statistical field theory: Volume 2 strong coupling, Monte Carlo Methods, conformal field theory, and random systems*. Cambridge: Cambridge University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.