Replica Analysis of the Linear Model with Markov or Hidden Markov Signal Priors

Lan V. Truong, *Member, IEEE*

***Abstract*—This paper estimates free energy, average mutual information, and minimum mean square error (MMSE) of a linear model under two assumptions: (1) the source is generated by a Markov chain, (2) the source is generated via a hidden Markov model. Our estimates are based on the replica method in statistical physics. We show that under the posterior mean estimator, the linear model with Markov sources or hidden Markov sources is decoupled into single-input AWGN channels with state information available at both encoder and decoder where the state distribution follows the left Perron-Frobenius eigenvector with unit Manhattan norm of the stochastic matrix of Markov chains. Numerical results show that the free energies and MSEs obtained via the replica method are closely approximate to their counterparts achieved by the Metropolis–Hastings algorithm or some well-known approximate message passing algorithms in the research literature.**

***Index Terms*—Compressed sensing, Linear model, Linear regression, Markov chain, Hidden Markov model, Replica method, Free energy, Minimum mean square error, Statistical Physics, Maximum a posteriori estimation.**

I. Introduction

In the canonical compressed sensing problem, the primary goal is to reconstruct an n-dimensional vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ with independent and identical prior from an m-dimensional vector of noisy linear observations $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_m)$ of the form $Y_k = \langle \boldsymbol{\Phi}_k, \boldsymbol{X} \rangle + W_k, k = 1, 2, \cdots, m$, where $\{\boldsymbol{\Phi}_k\}$ is a sequence of n-dimensional measurement vectors, $\{W_k\}$ is a sequence of standard Gaussian random variables, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product between vectors. In this paper, under the assumption that $\boldsymbol{X}$ has a Markov or hidden Markov prior, we wish to estimate the asymptotic mutual information $\lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y})$ and the MMSE $\lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{\Phi}]\|^2]$. Our estimates are based on the replica method which was developed originally to study mean field approximations in spin glasses [1]. Although this method lacks of rigorous mathematical proof in some particular parts, it has been widely accepted as an analytic tool and utilized to investigate a variety of problems in applied mathematics, information processing, machine learning, and coding [2].

A. Related Work

The use of the replica method for studying multiuser estimators goes back to [3] where Tanaka determined the asymptotic bit error rate of Marginal-Posterior-Mode (MPM) estimators by employing the replica method. The study demonstrated interesting large-system properties of multiuser estimators. As a result, the statistical physics approach received more attention in the context of multiuser systems [4], [5] with a subsequent work focusing on the compressed sensing directly [6]–[11]. Guo and Verdú [4] studied the same CDMA detection problem as [3] but under more general (arbitrary) input distributions. They assumed that a generic posterior mean estimator is applied before single-user decoding. The generic detector can be particularized to the matched filter, decorrelator, linear minimum mean-square error (MMSE) detector, the jointly or the individual optimal detector, and others. It is found that the detection output for each user, although in general asymptotically non-Gaussian conditioned on the transmitted symbol, converges as the number of users go to infinity to a deterministic function of a "hidden" Gaussian statistic independent of the interferers. Thus, the multi-user channel can be decoupled.

The results of replica method have been rigorously in a number of settings in compressed sensing. One example is given by message passing on matrices with special structure, such as sparsity [12]–[16] or spatial coupling [17]–[19]. In [8], Rangan et al. studied the asymptotic performance of a class of Maximize-A-Posterior (MAP) estimators. Using standard large deviation techniques, the authors represented the MAP estimator as the limit of an indexed MMSE estimator's sequence. Consequently, they determined the estimator's asymptotics employing the results from [4] and justified the decoupling property of MAP estimators under Replica Symmetry (RS) assumption for an i.i.d. measurement matrix $\boldsymbol{\Phi}$. The asymptotic performance for the MAP estimator where the RS assumption does not hold but satisfies some looser symmetric assumptions, called Replica Symmetry Breaking (RSB) is considered in [2]. Under the RSB assumption with b steps of breaking (bRSB), the equivalent noisy single-user channel is given in form of an input term added by an impairment term. The impairment term, moreover, is expressed as a sum of an independent Gaussian random variable and b correlated non-Gaussian interference terms.

Recently, there have been some works which aim to close the gap between mathematically rigorous proof and results from the replica method. Reeves and Pfister considered the fundamental limit of compressed sensing for i.i.d. signal distributions and i.i.d. Gaussian measurement matrices [20]. Under some mild technical conditions, their results show that the limiting mutual information and Minimum Mean Square Error (MMSE) are equal to the values predicted by the replica

L. Truong is with the Department of Engineering, University of Cambridge, United Kingdom (e-mail:lt407@cam.ac.uk).

This paper was presented in part at the 2021 International Symposium on Information Theory (ISIT) and the 25th International Conference on Artificial Intelligence and Statistics (AISTATS).

method. Their proof techniques are based on establishing relationships between mutual information and MMSE at finite n, m and $n \sim m$ such as [21], and extending obtained results in large system limits. In [22], Barbier et al. showed that the results for Generalized Linear Models (GLM) and i.i.d. sources stemming from the replica method are indeed correct and imply the optimal value of both estimation and generalization error. The proof is based on the adaptive interpolation method [23] which is an extension of interpolation method developed by Guerra and Toninelli [24] in the context of spin glasses, with an adaptive interpolation path. More specifically, this scheme interpolates between the original problem and the solution via replica method in small steps, each step involving its own set of trial parameters and Gaussian mean-fields in the spirit of Guerra and Toninelli. We are then able to choose the set of trial parameters in various ways so that the upper and lower bounds are eventually matched. By a generalization of the adaptive interpolation method, Truong [25] has recently established exact asymptotic expressions for the normalized mutual information and MMSE of sparse linear regression in the sub-linear sparsity regime, i.e., $m = n^\alpha$ for some $\alpha \in (0, 1)$. This work shows that the traditional linear assumption between the signal dimension and number of observations in the replica and adaptive interpolation methods is not necessary for sparse signals.

In all above research literature, the authors assume that the source is independently and identically distributed (i.i.d.). In many practical applications, samples of input data may be dependent on each other, e.g., Markov chains or hidden Markov models. There are a few non-rigorous literatures handling Markov chains using the replica method [26]–[28]. However, to the best of our knowledge, there exists no rigorously analytic result which was developed based on replica-related methods for these models. Some recent works considered the linear model with random generative priors where the signal is the output of a Bayesian neural network with specific structures with the input being an i.i.d. sequence [29]–[31]. Although these papers are to recover the structured signal, however, the signal structure is different from Markov or hidden Markov. For example, if we use a classifier (one layer neural network) with ReLU activation function, i.e., $\boldsymbol{x} = \sigma(\boldsymbol{a}^T \boldsymbol{u})$ where $\boldsymbol{a}$ is Gaussian as the assumptions in these papers and $\boldsymbol{u}$ is an i.i.d. vector, then $\boldsymbol{x}$ is not Markov or Hidden Markov. The adaptive interpolation method looks hard to apply for the linear model with Markov sources or hidden Markov sources since it requires that $X_1, X_2, \dots, X_n$ are i.i.d. (or at least i.i.d. block-by-block) to guarantee a fixed interpolating free energy at the final (k, t)-interpolation model for each finite value of n [23]. There were also some existing works related to Mean Square Errors (MSE) achieved by Approximate Message Passing algorithms (AMP) for the linear model with Markov or hidden Markov sources [32]–[34]. Approximate message passing (AMP) refers to a class of efficient algorithms for statistical estimation in high-dimensional problems such as compressed sensing and low-rank matrix estimation. AMP is initially proposed for sparse signal recovery and compressed sensing [35]–[37]. AMP algorithms have been proved to be effective in reconstructing sparse signals from a small number of incoherent linear measurements. Their dynamics are accurately tracked by a simple one-dimensional iteration termed *state evolution* [38]. The state evolution is redefined in non-asymptotic sense for the sparse linear regression with sublinear sparsity in [25]. AMP algorithms achieve state-of-the-art performance for several high-dimensional statistical estimation problems, including compressed sensing [18], [38], [39] and low-rank matrix estimation [38], [40].

B. Main Contributions

In this paper, based on the same replica assumptions as [4], we establish free energy, mutual information, and MMSE for the linear model with Markov or hidden Markov sources. When limiting to the linear model with i.i.d. sources as case, we recover Guo and Verdú's results [4], which extends Tanaka work [3] to more general alphabets. More specially, our main contributions are as follows:

- Using the replica method, we estimate the free energy, the normalized mutual information in the large system limit for two models: linear model with Markov sources and linear model with hidden Markov sources (cf. Claim 1 and Claim 3).
- Using the replica method, we characterize MMSEs in the large system limit for two estimation problems (cf. Claim 2 and Claim 3). We show that under the posterior mean estimator, the linear model with Markov sources or hidden Markov sources is decoupled into single-input AWGN channels with state information available at both encoder and decoder where the state distribution follows the left Perron-Frobenius eigenvector with unit Manhattan norm of the stochastic matrix of Markov chains[1].
- We show that the free energies and MSEs obtained via the replica method are closely approximate to their counterparts achieved by the Metropolis–Hastings algorithm or some well-known approximate message passing algorithms in the research literature (cf. Section IV).

Essentially, our results show that in the large system limit, we can convert the estimation in high-dimensional space for the linear model with Markov or hidden Markov signal prior to the estimation problems in one-dimensional spaces. Compared with the linear model with i.i.d. sources [4], we need to deal with some new technical challenges related to the estimation of the derivative of Perron-Frobenius eigenvalue of non-negative matrices. For example, in the following Lemma 7, we develop a new technique to estimate this derivative in the large system limit.

MMSE and free energy are very important fundamental limits, which are benchmarks to check if a coding scheme or a learning algorithm for the linear model is optimal. In this work, we aim to characterize these fundamental limits by using replica method. Our simulation results (cf. Section IV) show that some existing MCMC algorithms (for example, Metropolis–Hastings algorithm) and AMP (for example, Turbo AMP [32]) are (potentially) optimal for the linear model with

[1] For any irreducible Markov process $\{Z_n\}_{n=1}^\infty$, the left Perron-Frobenius eigenvector with unit Manhattan norm is the stationary distribution of this Markov process, and the Perron-Frobenius eigenvalue is equal to 1 [41].

Markov or hidden Markov signal prior. Before our work, whether these interesting algorithms are optimal or not is an open question.

C. Paper Organization

The problem setting is placed in Section II, where we introduce the system model, posterior mean estimation, free energy and replica method in statistical physics. We also introduce some new concepts such as single-symbol Posterior Mean Estimation (PME) channel with state information, free energy functions, and other related notations in this section. Our main results are stated and proved in Section III. We apply our main results to estimate free energy, mutual information, and MMSE for some specific Markov chains or hidden Markov models in Section IV, where we also compare our obtained MMSEs with achievable MSEs by the classical Metropolis–Hastings algorithm and some well-known AMP algorithms in research literature. Proofs of main results are placed in Section V.

D. Notation

Use $[n]$ to denote the set $\{1,\dots,n\}$. Random vectors and matrices are in bold letters. Expectations with respect to "quenched" random variables (i.e., the variables that are fixed by the realization of the problem) are denoted by $\mathbb{E}$ and those with respect to "annealed" random variables (i.e., dynamical variables) are denoted by Gibbs bracket $\langle - \rangle$ possibly with appropriate subscripts. This choice follows the stardards of statistical physics.

As standard literature, we define $x^n = (x_1, x_2, \cdots, x_n)^T$ to denote a vector of length n. However, if the dimension of a vector x is clear from context, we omit it for simplicity. Let $\log x := \log_2 x$ and $\ln x$ be the natural logarithm of x for all $x \in \mathbb{R}^+$. Manhattan and Euclidean norms of a vector $x \in \mathbb{R}^n$ are defined as

$$\|x\|_1 := \sum_{i=1}^n |x_i|, \tag{1}$$

$$\|x\|_2 := \sqrt{\sum_{i=1}^n |x_i|^2}, \tag{2}$$

respectively. In addition, $\mathrm{vec}(\cdot)$ denotes the vectorization operator. Besides, for any $A \in \mathbb{R}^{p\times q}$ and $B \in (\mathbb{R}^{n\times n})^{p\times q}$, we define $A o_{\mathrm{tr}} B := \sum_{i,j} A_{ij} \odot B_{ij}$, where $A \odot B$ is the Hadamard product between A and B.

The moment generating function of a random vector $\boldsymbol{X} \in \mathbb{R}^n$ is defined as $\mathcal{M}(\lambda) := \mathbb{E}[\exp(\lambda^T \boldsymbol{X})]$ for all $\lambda \in \mathbb{R}^n$. Let $\mathcal{M}(\tilde{Q}) := \mathbb{E}[\exp(\mathrm{tr}(\tilde{Q}\boldsymbol{Q}))]$ be the moment generating function of a random matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ for all matrix $\tilde{Q} \in \mathbb{R}^{n\times n}$.

Denote by

$$\mathcal{Q} := \left\{ s x x^T \ \text{ for some } \ x \in \mathcal{S} \times \mathcal{X}^{\nu+1} \right\}. \tag{3}$$

For simplicity of presentation, we enumerate all matrices in $\mathcal{Q}$ as $\bar{Q}_0, \bar{Q}_1, \cdots, \bar{Q}_M$ where $M := |\mathcal{Q}| - 1$.

II. Problem Setting

We consider the linear model

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{W} = \boldsymbol{A}\boldsymbol{S}^{1/2}\boldsymbol{X} + \boldsymbol{W}. \tag{4}$$

Here $\boldsymbol{Y} \in \mathbb{R}^m$ is a vector of observations, $\boldsymbol{X} \in \mathbb{R}^n$ is the signal vector, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a measurement matrix, $\boldsymbol{S}$ is diagonal matrix of positive scale factors:

$$\boldsymbol{S} = \mathrm{diag}(S_1, S_2, \cdots, S_n), \quad S_j \in \mathbb{R}^+, \tag{5}$$

and $\boldsymbol{W} \in \mathbb{R}^m$ is a noise vector. We consider a sequence of problems indexed by n, and make the following assumptions on the model. These assumptions are identical to those in earlier works [4], [8] except for the signal prior, which we allow to be Markov or hidden Markov in contrast to the i.i.d. priors considered in earlier works.

1) We assume that the number of measurements m scales linearly with n, and $\lim_{n\to\infty} \frac{n}{m} = \beta$, for some $\beta > 0$.
2) The elements $\{A_{ij}\}_{i\in[m], j\in[n]}$ of the matrix $\boldsymbol{A}$ are i.i.d. and distributed as $A_{ij} \overset{\mathrm{d}}{=} \frac{1}{\sqrt{m}} A$, where A is a random variable with zero mean, unit variance and all moments finite.
3) The scale factors $(S_1, \dots, S_n)$ are i.i.d. according to P_S, which is supported on a set $\mathcal{S} \subset \mathbb{R}^+$. The scale factors $(S_1, \dots, S_n)$ are independent of $\boldsymbol{A}, \boldsymbol{X}$, and $\boldsymbol{W}$.
4) The noise vector $\boldsymbol{W}$ is standard normal, i.e., $W_j \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$ for $j \in [m]$.
5) *Signal prior*: We assume that the components of $\boldsymbol{X}$ take values on a Polish space on $\mathbb{R}$, and are distributed according to either a Markov or a hidden Markov prior.
   - *Markov chain prior*: This model assumes that

     $$\begin{aligned}\mathbb{P}(\boldsymbol{X} = (x_1, \dots, x_n)) \\ = p(x_1)\pi(x_1, x_2)\cdots\pi(x_{n-1}, x_n)\end{aligned} \tag{6}$$

     for some initial probability distribution $p(\cdot)$ on $\mathcal{X}$, where $\pi(\cdot,\cdot)$ is the transition probability of a time-homogeneous, irreducible Markov chain on $\mathcal{X}$.
   - *Hidden Markov prior*: The second model assumes that $\{X_n\}_{n=1}^\infty$ are generated by a Hidden Markov Model (HMM), with hidden states $\{\Upsilon_n\}_{n=1}^\infty$ take values on a Polish space on $\mathcal{H}_\Upsilon$. That is, $\mathbb{P}(\boldsymbol{\Upsilon} = (\upsilon_1, \dots, \upsilon_n)) = p_\Upsilon(\upsilon_1)\pi_\Upsilon(\upsilon_1, \upsilon_2)\cdots\pi_\Upsilon(\upsilon_{n-1}, \upsilon_n)$ for some initial probability distribution $p_\Upsilon(\cdot)$ on $\mathcal{H}_\Upsilon$, where $\pi_\Upsilon(\cdot,\cdot)$ is the transition probability of a time-homogeneous, irreducible Markov chain on $\mathcal{H}_\Upsilon$. Then,

     $$\begin{aligned}\mathbb{P}(X_i = x_i \mid \Upsilon_1 = \upsilon_1, \dots, \Upsilon_i = \upsilon_i) \\ = p_{X|\Upsilon}(x_i \mid \upsilon_i), \quad i \in [n],\end{aligned}$$

     for some stationary emission probability $p_{X|\Upsilon}(\cdot|\cdot)$ on $\mathcal{S}_\Upsilon \times \mathcal{X}$.

For simplicity of presentation, we assume that Markov chains $\{X_n\}_{n=1}^\infty, \{\Upsilon_n\}_{n=1}^\infty$ have finite state spaces and $\mathcal{S}$ has a finite number of elements in some proofs. However, it is not hard to extend these proofs to Markov chains on Polish spaces in $\mathbb{R}$ with an infinite set $\mathcal{S}^2$ by referring to a more general definition of Markov chain in [42] and noting that the Varadhan's large deviation theorem holds for Markov chains on the general Polish space. An *irreducible* and *recurrent* Markov chain on an infinite state-space is called a Harris chain [42], which owns many similar properties to the finite state-space version such as the existence of an unique stationary distribution. For both models, we denote the joint probability mass distribution (pmf) of the signal by $p(x_1, \dots, x_n)$. For general proofs, we use Radon–Nikodym derivatives with respect to corresponding measures [43].

A. Posterior Mean Estimation

The problem setting described above induces a posterior distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Phi}}$, given by

$$p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\phi}) = \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\phi}) p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{\phi})}, \tag{7}$$

where

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \Phi) = (2\pi)^{-m/2} \exp\left[-\frac{\|\boldsymbol{y} - \boldsymbol{\phi}\boldsymbol{x}\|^2}{2}\right], \tag{8}$$

and

$$\begin{aligned} p_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{\phi}) &= \mathbb{E}_p[p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\phi})] \\ &= \sum_{\boldsymbol{x}} p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\phi}) p_{\boldsymbol{X}}(\boldsymbol{x}). \end{aligned} \tag{9}$$

The (canonical) posterior mean estimator (PME), which computes the mean value of the posterior distribution $p_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Phi}}$ is given by,

$$[\boldsymbol{X}] = \mathbb{E}_p\big[\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{\Phi}\big]. \tag{10}$$

This estimator achieves MMSE between the estimated and the original signal.

As in Guo and Verdú [4], we consider a more general class of posterior mean estimators, based on a *postulated* posterior distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Phi}}$, to model that scenario that the true posterior mean may be infeasible to compute or the estimator may not know the exact prior and the noise variance. The postulated posterior distribution is defined via a postulated prior and a postulated noise variance. The postulated prior $q_{\boldsymbol{X}}(\boldsymbol{x})$ is of the form

$$q_{\boldsymbol{X}}(\boldsymbol{x}) = q(x_1)\tilde{\pi}(x_1, x_2) \cdots \tilde{\pi}(x_{n-1}, x_n) \tag{11}$$

for some initial distribution $q(\cdot)$ on $\mathcal{X}$, and $\tilde{\pi}(\cdot, \cdot)$ is the transition probability of an irreducible Markov chain on $\mathcal{X}$. The postulated likelihood is Gaussian with variance σ^2, which may not be equal to the true noise variance 1:

$$q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\phi}) = (2\pi)^{-m/2} \exp\left[-\frac{\|\boldsymbol{y} - \boldsymbol{\phi}\boldsymbol{x}\|^2}{2\sigma^2}\right] \tag{12}$$

The postulated prior and noise variance induce the posterior distribution $q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Phi}}$ given by

$$q_{\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\phi}) = \frac{q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\phi})\, q_{\boldsymbol{X}}(\boldsymbol{x})}{q_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{y}|\boldsymbol{\phi})}, \tag{13}$$

where

$$\begin{aligned} q_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{y}|\boldsymbol{\phi}) &= \mathbb{E}_q[q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\phi}) \mid \boldsymbol{\Phi} = \boldsymbol{\phi}] \\ &= \sum_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x})\, q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\phi}). \end{aligned} \tag{14}$$

The posterior mean estimator computed using (13), which we call the 'generalized PME', is denoted by

$$[\boldsymbol{X}]_q = \mathbb{E}_q\big[\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{\Phi}\big]. \tag{15}$$

As described in [4], with suitable choices of the postulated distribution, the generalized PME can recover various commonly used sub-optimal estimators such as the linear MMSE estimator and the matched filter. The postulated prior can also be used to model estimators that ignore the memory in the signal $\boldsymbol{X}$, e.g., estimator based on an i.i.d. prior.

In the remainder of the paper, we will use the subscript p to denote expectations computed using the true prior/posterior, and q to denote expectations using the postulated prior/posterior.

B. Free Energy and Replica Method

Let

$$Z(\boldsymbol{Y}, \boldsymbol{\Phi}) := q_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{Y}|\boldsymbol{\Phi}). \tag{16}$$

The free energy of the model in (4) is defined as

$$\mathcal{F}_n := -\frac{1}{n} \log Z(\boldsymbol{Y}, \boldsymbol{\Phi}). \tag{17}$$

The expectation of the free energy (with respect to $q_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{Y}|\boldsymbol{\Phi})$) is equal to the conditional entropy of the observation $\frac{1}{n}H_q(\boldsymbol{Y}|\boldsymbol{\Phi})$ as well as (up to an additive constant) to the mutual information density between the signal and the observations $\frac{1}{n}I_q(\boldsymbol{X}, \boldsymbol{Y})$.

The asymptotic free energy is the limit of the sequence $\{F_n\}_{n=1}^{\infty}$, i.e.,

$$\mathcal{F}_q := \lim_{n\to\infty} \mathcal{F}_n. \tag{18}$$

In general, it is very challenging to prove the existence and estimate the limit in (18). Replica method, originally developed in statistical physics, is usually used to evaluate this limit [3], [4] because the linear model is similar to the thermodynamic system. For this model, replica method is based on the following assumptions (A) and facts (F):

- (A1) The free energy $\mathcal{F}_n$ has the self-averaging property as $n \to \infty$. This means that

$$\mathcal{F} := \lim_{n\to\infty} \mathbb{E}[\mathcal{F}_n]. \tag{19}$$

  The self-averaging property essentially assumes that the variations of $Z(\boldsymbol{Y}, \boldsymbol{\Phi})$ due to the randomness of the measurement matrix Φ vanish in the limit $n \to \infty$. Although a large number of statistical physics quantities exhibit such self-averaging, the self-averaging of the relevant quantities for the general PME (PMMSE) and Postulated MAP (PMAP) analyses has not been rigorously established [8]. For the purpose of estimating the average mutual information of the Markov model only, we don't need to make use of this assumption.

[2] In the Appendix B, we show how to extend our analysis to Markov chain on a general Polish spaces in $\mathbb{R}$.

- (F1) The following identity holds:
$$\mathbb{E}[\log Z(\boldsymbol{Y},\boldsymbol{\Phi})] = \lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\mathbb{E}[Z^{\nu}(\boldsymbol{Y},\boldsymbol{\Phi})]. \quad (20)$$
- (A2) Estimation of $\mathbb{E}[Z(\boldsymbol{Y},\boldsymbol{\Phi})^{\nu}]$ for a positive real number ν in the neighbourhood of 0 can be done by two steps: (1) Estimate $\mathbb{E}[Z^{\nu}(\boldsymbol{Y},\boldsymbol{\Phi})]$ for a general positive integer ν (2) Take the limit of the obtained result as $\nu\to 0$. This is called "replica trick" in statistical physics.
- (F2) For any positive integer ν and a realization $(\boldsymbol{y},\Phi)$ of $(\boldsymbol{Y},\boldsymbol{\Phi})$, the quantity $Z^{\nu}(\boldsymbol{y},\boldsymbol{\Phi})$ can be written as
$$Z^{\nu}(\boldsymbol{y},\Phi) = \Big\{q_{\boldsymbol{Y}|\boldsymbol{\Phi}}(\boldsymbol{y}|\Phi)\Big\}^{\nu} \quad (21)$$
$$= \Big\{\mathbb{E}_{q_{\boldsymbol{X}}}\Big[q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y}|\boldsymbol{X},\Phi)\Big]\Big\}^{\nu} \quad (22)$$
$$= \mathbb{E}_{q_{\boldsymbol{X}}}\Big\{\prod_{a=1}^{\nu} q_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Phi}}(\boldsymbol{y}|\boldsymbol{X}^{(a)},\Phi)\Big\}. \quad (23)$$
where the last expectation is taken over relicated vectors $\boldsymbol{X}^{(a)}, a=1,2,\cdots,\nu$ which are independent copies of a random vector with postulated distribution $q_{\boldsymbol{X}}$.
- (A3) The order of limit $n\to\infty$ and $\nu\to 0$ can be interchanged. Mathematically, under some conditions such as Theorem Moore-Osgood [44], the interchange between limits work. This theorem is used in [45] for a similar purpose.
- (A4) Usually, the free energy can be expressed an optimal value of an optimization problem over the space of covariance matrices of replica samples, say $\mathcal{Q}$. This optimization is general difficult to perform. To overcome this, the replica method also makes an additional assumption that the optimizer Q^* is symmetric with respect to permutations of ν replica indices. This assumption is called Replica Symmetry (RS) in statistical physics. See Definition 10 for our assumption about RS in this paper.

III. Main results

A. Results for Markov Priors

Our results on the free energy and MMSE will be stated in terms of a *single-symbol* channel, similar to the equivalent single-user Gaussian channel which is obtained via decoupling as in [4, Section D]. Let $\lambda^{(\pi)}$ be the left Perron-Frobenius eigenvector with unit Manhattan norm[3] of $P_\pi = \{\pi(x,y)\}_{x\in\mathcal{X},y\in\mathcal{X}}$ which is the stochastic matrix of the Markov chain $\{X_n\}_{n=1}^{\infty}$, and let $\lambda_{x_0}^{(\pi)}$ be the component of $\lambda^{(\pi)}$ associated with the x_0-th row of P_π. Let us consider the composition of a Gaussian channel with one state X_0 available at both encoder and decoder such that $\mathsf{X}_0 = x_0 \sim \lambda_{x_0}^{(\pi)}$, a one-state PME, and a companion retrochannel in the single-symbol setting depicted in Fig. 1. Given the state information $\mathsf{X}_0 = x_0\sim\lambda_{x_0}^{(\pi)}$, the input-output relationship of this single-symbol channel is given by
$$U = \sqrt{S}\,\mathsf{X}_1 + \frac{1}{\sqrt{\eta}}W, \quad (24)$$
where the input $\mathsf{X}_1 \sim p_{\mathsf{X}_1|\mathsf{X}_0}(\cdot|x_0) := \pi(x_0,\cdot)$, $S\sim P_S$ which is independent X_0 and X_1, $W\sim\mathcal{N}(0,1)$ the noise independent of X_0 and X_1, and $\eta>0$ the inverse noise variance. The conditional distribution associated with the channel is
$$p_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u\mid x_0,x_1,s;\eta) = \sqrt{\frac{\eta}{2\pi}}\exp\Big[-\frac{\eta}{2}(u-\sqrt{s}x_1)^2\Big]. \quad (25)$$
Let $q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\xi}$ represent Gaussian channel with state X_0 available at both encoder and decoder akin to (24), the only difference being that the inverse noise variance is ξ instead of η
$$q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\xi}(u\mid x_0,x_1,s;\xi) = \sqrt{\frac{\xi}{2\pi}}\exp\Big[-\frac{\xi}{2}(u-\sqrt{s}x_1)^2\Big]. \quad (26)$$
Similar to that in the vector channel setting, by postulating the input distribution to be $q_{\mathsf{X}_1|\mathsf{X}_0}(\cdot|x_0)=\tilde{\pi}(x_0,\cdot)$, a posterior probability distribution $q_{\mathsf{X}_1|\mathsf{X}_0,U,S;\xi}$ is induced by $q_{\mathsf{X}_1|\mathsf{X}_0}$ and $q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\xi}$ using the Bayes rule, i.e.,
$$q_{\mathsf{X}_1|\mathsf{X}_0,S,U;\xi}(x\mid x_0,s,u;\xi) = \frac{q_{\mathsf{X}_1|\mathsf{X}_0}(x\mid x_0)q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\xi}(u\mid x_0,x_1,s;\xi)}{q_{U|\mathsf{X}_0,S;\xi}(u\mid x_0,s;\xi)}. \quad (27)$$
This induces a single-use retrochannel with random transformation $q_{\mathsf{X}_1|\mathsf{X}_0,U,S;\xi}$, which outputs a random variable X given the channel output U and the channel state X_0 (Fig. 1). A (generalized) single-symbol PME with state available $\mathsf{X}_0=x_0$ is defined naturally as (cf. (15))
$$\langle\mathsf{X}|\mathsf{X}_0=x_0\rangle_q = \mathbb{E}_q\big[\mathsf{X}|\mathsf{X}_0=x_0,U,S;\xi\big], \quad (28)$$
where the expectation is taken over the (conditionally) postulated distribution in (27).

The single-symbol PME (28) is merely a decision function applied to the Gaussian channel output with state $\mathsf{X}_0=x_0$ available at both encoder and decoder (or input and output), which can be expressed explicitly as
$$\mathbb{E}_q\big[\mathsf{X}|U,\mathsf{X}_0=x_0,S;\xi\big] = \frac{q_1(U,x_0,S;\xi)}{q_0(U,x_0,S;\xi)}, \quad (29)$$
where
$$q_0(u,x_0,S;\xi) := q_{U|\mathsf{X}_0,S;\xi}(u\mid x_0,S;\xi) = \mathbb{E}\Big[q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\xi}(u\mid x_0,\mathsf{X}_1,S;\xi)\Big|S\Big], \quad (30)$$
$$q_1(z,x_0,S;\xi) = \mathbb{E}\Big[\mathsf{X}\,q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\xi}(z\mid x_0,\mathsf{X}_1,S;\xi)\Big|S\Big]. \quad (31)$$

The probability law of the (composite) single-symbol channel depicted by Fig. 1 is determined by S and two parameters η and ξ given state X_0. We define the conditional mean-square error of the PME as
$$\mathcal{E}(S;\eta,\xi|x_0) = \mathbb{E}[(\mathsf{X}_1-\langle\mathsf{X}\,|\,\mathsf{X}_0\rangle_q)^2\mid\mathsf{X}_0=x_0,S;\eta,\xi] \quad (32)$$
and also define the conditional variance of the retrochannel as
$$\mathcal{V}(S;\eta,\xi|x_0) = \mathbb{E}\big[(\mathsf{X}-\langle\mathsf{X}|\mathsf{X}_0\rangle_q)^2\mid\mathsf{X}_0=x_0,S;\eta,\xi\big]. \quad (33)$$

[3]Since there exists a unique left Perron-Frobenius eigenvector up to a positive scaling factor [41], $\lambda^{(\pi)}$ exists uniquely, which is the stationary distribution of the Markov chain.

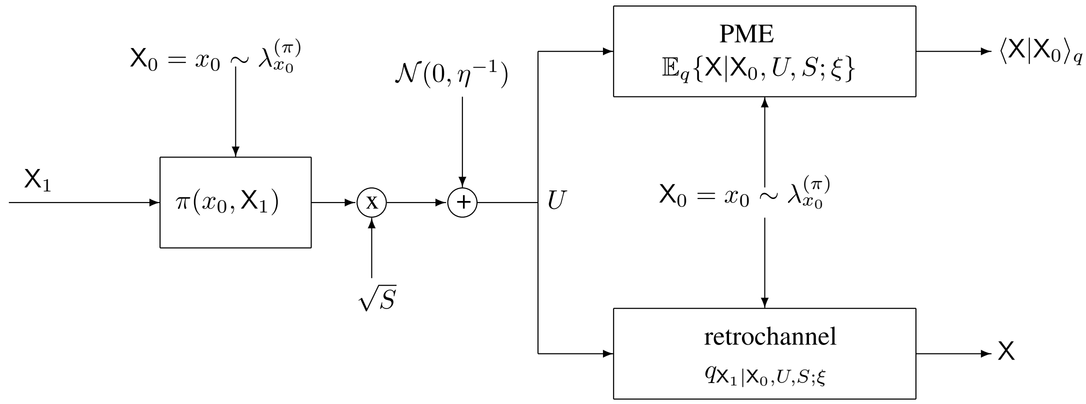

Fig. 1. The equivalent single-symbol Gaussian channel with state available at both encoder and decoder, PME, and retrochannel.

Define

$$\mathcal{G} := \sum_{x_0\in\mathcal{X}} \lambda_{x_0}^{(\pi)}\mathcal{G}(x_0), \tag{34}$$

where

$$\begin{aligned}\mathcal{G}(x_0)&\\ := -\mathbb{E}\Big\{ &\int p_{U|X_0,X_1,S;\eta}(u \mid x_0, X_1, S;\eta)\\ &\quad\times \log q_{U|X_0,X_1,S;\xi}(u \mid x_0, S;\xi)du\Big\}\\ &+\frac{1}{2\beta}\Big[(\xi-1)\log e - \log\xi\Big] - \frac{1}{2}\log\frac{2\pi}{\xi} - \frac{\xi}{2\eta}\log e\\ &+\frac{\sigma^2\xi(\eta-\xi)}{2\beta\eta}\log e + \frac{1}{2\beta}\log(2\pi) + \frac{\xi}{2\beta\eta}\log e,\end{aligned} \tag{35}$$

and η and ξ is the solution of the following equation system

$$\eta^{-1} = 1+\beta\sum_{x_0\in\mathcal{X}}\lambda_{x_0}^{(\pi)}\mathbb{E}[S\mathcal{E}(S;\eta,\xi \mid x_0)], \tag{36}$$

$$\xi^{-1} = \sigma^2+\beta\sum_{x_0\in\mathcal{X}}\lambda_{x_0}^{(\pi)}\mathbb{E}[S\mathcal{V}(S;\eta,\xi \mid x_0)] \tag{37}$$

such that they minimize $\mathcal{G}$. Observe that for the case $X_0, X_1, \cdots, X_n$ are i.i.d., $\mathcal{G}(x_0)$ does not depend on x_0 and is defined in [4, Eq. (22)].

Claim 1. *The free energy of the linear model with Markov sources in Section II satisfies*

$$\mathcal{F}_q = \mathcal{G}, \tag{38}$$

where $\mathcal{G}$ is defined in (34). In addition, the average mutual information of this model satisfies:

$$C = \lim_{n\to\infty}\frac{1}{n}I\big(\boldsymbol{X}^n;\boldsymbol{Y}^m|\boldsymbol{\Phi}\big) = \mathcal{F}_q\Big|_{\sigma=1} - \frac{1}{2\beta}. \tag{39}$$

Claim 2. *Recall the definition of $\{\lambda_{x_0}^{(\pi)}\}_{x_0\in\mathcal{X}}$ in Section III-A. Assume that the generalized PME defined in (15) is used for estimation. Then, for all $k\in[n]$ and $(i_0,j_0,l_0)\in\mathbb{Z}_+\times\mathbb{Z}_+\times\mathbb{Z}_+$, the joint moments satisfy:*

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{E}\big[X_k^{i_0}\tilde{X}_k^{j_0}[X_k]_q^{l_0}\big]\\ &\quad= \sum_{x_0\in\mathcal{X}}\lambda_{x_0}^{(\pi)}\mathbb{E}\big[X_1^{i_0}X^{j_0}\langle X|X_0\rangle_q^{l_0}\big|X_0=x_0\big],\end{aligned} \tag{40}$$

where $(X_1, X_0, X, \langle X|X_0\rangle_q)$ is the input, channel state, and outputs defined in the (composite) single-symbol PME channel in Fig. 1, and $(X_k,\tilde{X}_k,[X_k]_q)$ is the k-th symbol in the vector $\boldsymbol{X}\in\mathcal{X}^n$, the k-th output of the vector retrochanel defined in (13), and its corresponding estimated symbol by using the PME estimate in (15).

In addition, the average MMSE satisfies:

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\big[\|\boldsymbol{X}-[\boldsymbol{X}]\|_2^2\big]\\ &\quad=\mathbb{E}\big[X_1^2\big]-\sum_{x_0\in\mathcal{X}}\lambda_{x_0}^{(\pi)}\mathbb{E}\big[\langle X|X_0=x_0\rangle^2\big],\end{aligned} \tag{41}$$

where $X_1\sim\sum_{x_0\in\mathcal{X}}\pi(x_0,\cdot)\lambda_{x_0}^{(\pi)}$.

Observe that an i. i. d. sequence $\{X_n\}_{n=1}^\infty$ can be considered as a Markov sequence with transition probability (function) $\pi(x,y) = p(y)$ for all $x,y\in\mathcal{X}$. Hence, Claim 1 and Claim 2 can recover all results for the linear model with i.i.d. signal prior in [4, Sect. II-D]. For this special case, $\mathcal{G}(x_0)$ is a constant, say $\mathcal{G}(\emptyset)$, for all $x_0\in\mathcal{X}$. Here, $\mathcal{G}(\emptyset)$ is the free energy function estimated in Section III-A when there is no state information appeared in the correponding single-symbol PME channel, i.e. $X_0=\emptyset$. In addition, the left Perron-Frobenius eigenvector with unit Manhattan norm for the stochastic matrix for this special case is $\{P_{X_1}(x)\}_{x\in\mathcal{X}}$.

B. Results for Hidden Markov Priors

As the previous section, for the case that $\{X_n\}_{n=1}^\infty$ are hidden states of a Markov chain $\{\Upsilon_n\}_{n=1}^\infty$ on the space $\mathcal{S}_\Upsilon$, we define a new single-symbol channel with state which is similar to the conditional PME channel defined in Section

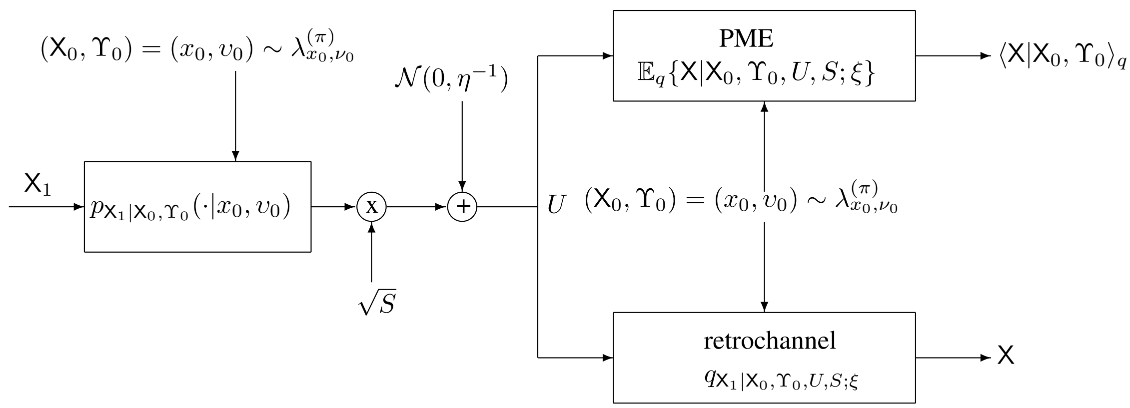

Fig. 2. The equivalent single-symbol Gaussian channel with two states available at both encoder and decoder, PME, and retrochannel.

III-A. Let $\lambda^{(\pi_\Upsilon)}$ be the left Perron-Frobenius eigenvector with unit Manhattan norm[4] of

$$P_{\pi,\mathcal{X}} = \Big\{P_{\mathsf{X}_1,\Upsilon_1|\mathsf{X}_0,\Upsilon_0}(x_1, \upsilon_1 \mid x_0, \upsilon_0)\Big\}_{(x_0,\upsilon_0),(x_1,\upsilon_1)\in\mathcal{X}\times\mathcal{S}_\Upsilon},$$
$$= \Big\{\pi_\Upsilon(\upsilon_0, \upsilon_1)P_{\mathsf{X}|\Upsilon}(x_1 \mid \upsilon_1)\Big\}_{(x_0,\upsilon_0),(x_1,\upsilon_1)\in\mathcal{X}\times\mathcal{S}_\Upsilon},$$

which is the stochastic matrix of the Markov chain $\{(X_n, \Upsilon_n)\}_{n=1}^{\infty}$[5], and let $\lambda^{(\pi)}_{x_0,\upsilon_0}$ be the component of $\lambda^{(\pi)}$ associated with the (x_0, υ_0)-th row of $P_{\pi,\mathcal{X}}$. Let us consider the composition of a Gaussian channel with two states $(\mathsf{X}_0, \Upsilon_0)$ available at both encoder and decoder such that $(\mathsf{X}_0, \Upsilon_0) = (x_0, \upsilon_0) \sim \lambda^{(\pi)}_{x_0,\upsilon_0}$, a two-state PME, and a companion retrochannel in the single-symbol setting depicted in Fig. 2. Given the state information $(\mathsf{X}_0, \Upsilon_0) = (x_0, \upsilon_0)$, the input-output relationship of this single-symbol channel is given by

$$U = \sqrt{S}\mathsf{X}_1 + \frac{1}{\sqrt{\eta}}W, \tag{42}$$

where the input $\mathsf{X}_1 \sim p_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0}(\cdot|x_0, \upsilon_0)$ such that

$$\begin{aligned} &p_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0}(x_1 \mid x_0, \upsilon_0) \\ &\quad= \sum_{\upsilon_1\in\mathcal{S}_\Upsilon} p_{\mathsf{X}_1,\Upsilon_1|\mathsf{X}_0,\Upsilon_0}(x_1, \upsilon_1 \mid x_0, \upsilon_0) &(43)\\ &\quad= \sum_{\upsilon_1\in\mathcal{S}_\Upsilon} \pi_\Upsilon(\upsilon_0, \upsilon_1)p_{\mathsf{X}|\Upsilon}(x_1 \mid \upsilon_1), &(44)\end{aligned}$$

$S \sim P_S$ which is independent X_0, Υ_0 and X_1, Υ_1, $W \sim \mathcal{N}(0, 1)$ the noise independent of X_0, Υ_0 and X_1, Υ_1, and $\eta > 0$ the inverse noise variance. The conditional distribution associated with the channel is

$$\begin{aligned} &p_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\eta}(u \mid x_0, \upsilon_0, x_1, s; \eta) \\ &\quad= \sqrt{\frac{\eta}{2\pi}} \exp\Big[-\frac{\eta}{2}(u - \sqrt{s}x_1)^2\Big]. \end{aligned} \tag{45}$$

[4]Since there exists a unique left Perron-Frobenius eigenvector unique up to a positive scaling factor [41], so $\lambda^{(\pi_\Upsilon)}$ exists uniquely.

[5]The fact that $\{(X_n, \Upsilon_n)\}_{n=1}^{\infty}$ forms a Markov chain can be easily proved.

Let $q_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\xi}$ represent Gaussian channel with two states X_0 and Υ_0 available at both encoder and decoder akin to (42), the only difference being that the inverse noise variance is ξ instead of η

$$\begin{aligned} &q_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\xi}(u \mid x_0, \upsilon_0, x_1, s; \eta, x_0) \\ &\quad= \sqrt{\frac{\xi}{2\pi}} \exp\Big[-\frac{\xi}{2}(u - \sqrt{s}x_1)^2\Big]. \end{aligned} \tag{46}$$

Similar to that in the vector channel setting, by postulating the input distribution to be $q_{\Upsilon_1|\Upsilon_0}(\cdot|\upsilon_0) = \tilde{\pi}_\Upsilon(\upsilon_0, \cdot)$, a posterior probability distribution $q_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0,U,S;\xi}$ is induced by $q_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0}$ and $q_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\xi}$ using the Bayes rule, i.e.,

$$\begin{aligned} &q_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0,S,U;\xi}(x \mid x_0, \upsilon_0, s, u; \xi) \\ &\quad= \frac{q_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0}(x \mid x_0, \upsilon_0)q_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\xi}(u \mid x_0, \upsilon_0, x_1, s; \xi)}{q_{U|\mathsf{X}_0,\Upsilon_0,S;\xi}(u \mid x_0, \upsilon_0, s; \xi)}. \end{aligned} \tag{47}$$

This induces a single-use retrochannel with random transformation $q_{\mathsf{X}_1|\mathsf{X}_0,\Upsilon_0,U,S;\xi}$, which outputs a random variable X given the channel output U and the channel states X_0, Υ_0 (Fig. 2). A (generalized) single-symbol PME with two available states $\mathsf{X}_0 = x_0$ and $\Upsilon_0 = \upsilon_0$ is defined naturally as (cf. (28))

$$\begin{aligned} &\langle \mathsf{X}|\mathsf{X}_0 = x_0, \Upsilon = \upsilon_0\rangle_q \\ &\quad= \mathbb{E}_q\big[\mathsf{X}|\mathsf{X}_0 = x_0, \Upsilon_0 = \upsilon_0, U, S; \xi\big], \end{aligned} \tag{48}$$

where the expectation is taken over the (conditionally) postulated distribution in (27).

The single-symbol PME (28) is merely a decision function applied to the Gaussian channel output with two states $\mathsf{X}_0 = x_0$ and $\Upsilon_0 = \upsilon_0$ available at both encoder and decoder (or input and output), which can expressed explicitly as

$$\begin{aligned} &\mathbb{E}_q\big[\mathsf{X}|U, \mathsf{X}_0 = x_0, \Upsilon_0 = \upsilon_0, S; \xi\big] \\ &\quad= \frac{q_1(U, x_0, \upsilon_0, S; \xi)}{q_0(U, x_0, \upsilon_0, S; \xi)}, \end{aligned} \tag{49}$$

where

$$q_0(u,x_0,\upsilon_0,S;\xi) := q_{U|\mathsf{X}_0,\Upsilon_0,S;\xi}(u \mid x_0,\upsilon_0,S;\xi) = \mathbb{E}\big[q_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\xi}(u \mid x_0,\upsilon_0,\mathsf{X}_1,S;\xi)\big|S\big], \tag{50}$$

$$q_1(u,x_0,\upsilon_0,S;\xi) = \mathbb{E}\big[\mathsf{X} q_{U|\mathsf{X}_0,\Upsilon_0,\mathsf{X}_1,S;\xi}(u \mid x_0,\upsilon_0,\mathsf{X}_1,S;\xi)\big|S\big]. \tag{51}$$

The probability law of the (composite) single-symbol channel depicted by Fig. 2 is determined by S and two parameters η and ξ given states X_0 and Υ_0. We define the conditional mean-square error of the PME as

$$\mathcal{E}(S;\eta,\xi|x_0,\upsilon_0) = \mathbb{E}[(\mathsf{X}_1 - \langle \mathsf{X}|\mathsf{X}_0,\Upsilon_0\rangle_q)^2|\mathsf{X}_0 = x_0, \Upsilon_0=\upsilon_0, S;\eta,\xi] \tag{52}$$

and also define the conditional variance of the retrochannel as

$$\mathcal{V}(S;\eta,\xi|x_0,\upsilon_0) = \mathbb{E}\big[(\mathsf{X} - \langle \mathsf{X}|\mathsf{X}_0,\Upsilon_0\rangle_q)^2|\mathsf{X}_0 = x_0, \Upsilon_0=\upsilon_0, S;\eta,\xi\big]. \tag{53}$$

Define

$$\tilde{G} := \sum_{(x_0,\upsilon_0)\in\mathcal{X}\times\mathcal{S}_\Upsilon} \lambda^{(\pi_\Upsilon)}_{x_0,\upsilon_0}\tilde{G}(x_0,\upsilon_0), \tag{54}$$

where

$$\begin{aligned}\tilde{G}(x_0,\upsilon_0) &:= -\mathbb{E}\Big\{\int p_{U|\mathsf{X}_0,\Upsilon_0,S;\eta}(u \mid x_0,\upsilon_0,S;\eta) \\ &\quad\times \log q_{U|\mathsf{X}_0,\Upsilon_0,S;\xi}(u \mid x_0,\upsilon_0,S;\xi)du\Big\} \\ &\quad+\frac{1}{2\beta}\Big[(\xi-1)\log e - \log\xi\Big] - \frac{1}{2}\log\frac{2\pi}{\xi} - \frac{\xi}{2\eta}\log e \\ &\quad+\frac{\sigma^2\xi(\eta-\xi)}{2\beta\eta}\log e + \frac{1}{2\beta}\log(2\pi) + \frac{\xi}{2\beta\eta}\log e,\end{aligned} \tag{55}$$

and η and ξ is the solution of the following equation system

$$\eta^{-1} = 1 + \beta\sum_{x_0,\upsilon_0}\lambda^{(\pi_\Upsilon)}_{x_0,\upsilon_0}\mathbb{E}[S\mathcal{E}(S;\eta,\xi \mid x_0,\upsilon_0)], \tag{56}$$

$$\xi^{-1} = \sigma^2 + \beta\sum_{x_0,\upsilon_0}\lambda^{(\pi_\Upsilon)}_{x_0,\upsilon_0}\mathbb{E}[S\mathcal{V}(S;\eta,\xi \mid x_0,\upsilon_0)] \tag{57}$$

such that they minimize $\tilde{G}$.

Claim 3. *Assume that $\{X_n\}_{n=1}^\infty$ is the hidden states (outputs) of a hidden Markov model generated by a Markov chain $\{\Upsilon_n\}_{n=1}^\infty$ with transition probability (function) $\pi_\Upsilon(\cdot,\cdot)$ on some Polish space $\mathcal{S}_\Upsilon$, i.e.,*

- *Υ_n is a Markov process and is not directly observable.*
- *$\mathbb{P}(X_n\in\mathcal{A}|\Upsilon_1=\upsilon_1,\Upsilon_2=\upsilon_2,\cdots,\Upsilon_n=\upsilon_n) = \mathbb{P}(X_n\in\mathcal{A}|\Upsilon_n=\upsilon_n) = P_{\mathsf{X}|\Upsilon}(\mathcal{A}|\upsilon_n)$,*

for every $n\geq 1$, $\upsilon_1,\upsilon_2,\cdots,\upsilon_n$, and an arbitrary measurable set $\mathcal{A}$, where $P_{\mathsf{X}|\Upsilon}(\cdot|\cdot)$ is some probability measure called emission probability. Then, the following holds:

- *$\{X_n,\Upsilon_n\}_{n=1}^\infty$ forms a Markov chain on $\mathcal{X}\times\mathcal{S}_\Upsilon$ with transition probability $P_{\mathsf{X}_1,\Upsilon_1|\mathsf{X}_0,\Upsilon_0}(x_1,\upsilon_1|x_0,\upsilon_0) = P_{\mathsf{X}|\Upsilon}(x_1|\upsilon_1)\pi_\Upsilon(\upsilon_0,\upsilon_1)$.*
- *Recall the definitions of $\{\lambda^{(\pi_\Upsilon)}_{x_0,\upsilon_0}\}_{(x_0,\upsilon_0)\in\mathcal{X}\times\mathcal{S}_\Upsilon}$ and $\tilde{G}$ in (55). Then, the free energy, mutual information, joint moments, the average MMSE of the linear model with hidden Markov sources in II satisfy:*

$$\mathcal{F}_q = \tilde{G}, \tag{58}$$

$$C = \mathcal{F}_q\Big|_{\sigma=1} - \frac{1}{2\beta}, \tag{59}$$

$$\lim_{n\to\infty}\mathbb{E}\big[X_k^{i_0}\tilde{X}_k^{j_0}[X_k]_q^{l_0}\big] = \sum_{x_0,\upsilon_0}\lambda^{(\pi_\Upsilon)}_{x_0,\upsilon_0}\mathbb{E}\big[\mathsf{X}_1^{i_0}\mathsf{X}^{j_0}\langle\mathsf{X}|\mathsf{X}_0,\Upsilon_0\rangle_q^{l_0}\big|\mathsf{X}_0=x_0,\Upsilon_0=\upsilon_0\big], \quad \forall i_0,j_0,l_0\in\mathbb{Z}_+, \tag{60}$$

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\|\boldsymbol{X}-[\boldsymbol{X}]\|_2^2] = \mathbb{E}[\mathsf{X}_1^2] - \sum_{x_0,\upsilon_0}\lambda^{(\pi_\Upsilon)}_{x_0,\upsilon_0}\mathbb{E}[\langle\mathsf{X}|\mathsf{X}_0=x_0,\Upsilon_0=\upsilon_0\rangle^2], \tag{61}$$

where $(\mathsf{X}_1,\mathsf{X},\langle\mathsf{X}|\mathsf{X}_0=x_0,\Upsilon_0=\upsilon_0\rangle_q)$ is the input and outputs defined in the (composite) single-symbol PME channel in Fig. 2, and $(X_k,\tilde{X}_k,[X_k]_q)$ is the k-th symbol in the vector $\boldsymbol{X}\in\mathcal{X}^n$, the k-th output of the vector retrochanel defined in (13), and its corresponding estimated symbol by using the generalized PME estimate in (15). In addition, in (61), $\mathsf{X}_1\sim\sum_{\upsilon\in\mathcal{S}_\Upsilon}P_{\mathsf{X}|\Upsilon}(\cdot|\upsilon)\pi_\Upsilon(\upsilon_0,\upsilon)$, where $P_{\mathsf{X}|\Upsilon}$ is the stationary emission probability of the hidden Markov process.

IV. Numerical Examples and Comparison with Algorithmic Performance

A. Binary-valued Markov Prior

Assume that $\boldsymbol{X}$ is a homogeneous Markov chain on the alphabet $\mathcal{X}=\{-1,1\}$ with the stochastic matrix as follows:

$$P_\pi = \begin{bmatrix}\pi(-1,-1) & \pi(-1,1)\\ \pi(1,-1) & \pi(1,1)\end{bmatrix} = \begin{bmatrix}1-\alpha & \alpha\\ \delta & 1-\delta\end{bmatrix} \tag{62}$$

for some α and δ in $(0,1)$.

1) Free Energy and Average Mutual Information: It is easy to see that the left Perron-Frobenius eigenvector $\lambda^{(\pi)}$ of P_π defined in Subsection III-A is

$$\lambda^{(\pi)} = \left(\frac{\delta}{\alpha+\delta},\frac{\alpha}{\alpha+\delta}\right)^T. \tag{63}$$

We assume that all postulated distributions are the same as their true ones for simplicity. We also assume that $S=1$ with probability 1. For this case, $\eta=\xi$, and (36)–(37) is degraded to the following equation in η:

$$\eta^{-1} = 1+\beta\bigg(\frac{\delta}{\delta+\alpha}\mathbb{E}_{f_U^{(1)}}\bigg[1-\bigg(\frac{1-\big(\frac{1-\alpha}{\alpha}\big)\exp(-2\eta U)}{1+\big(\frac{1-\alpha}{\alpha}\big)\exp(-2\eta U)}\bigg)^2\bigg] + \frac{\alpha}{\delta+\alpha}\mathbb{E}_{f_U^{(2)}}\bigg[1-\bigg(\frac{1-\big(\frac{\delta}{1-\delta}\big)\exp(-2\eta U)}{1+\big(\frac{\delta}{1-\delta}\big)\exp(-2\eta U)}\bigg)^2\bigg]\bigg), \tag{64}$$

where

$$f_U^{(1)}(u) = \frac{1}{\sqrt{2\pi}}(1-\alpha)\sqrt{\eta}\exp\left(-\frac{(u+1)^2\eta}{2}\right) + \frac{1}{\sqrt{2\pi}}\alpha\sqrt{\eta}\exp\left(-\frac{(u-1)^2\eta}{2}\right) \tag{65}$$

and

$$f_U^{(2)}(u) = \frac{1}{\sqrt{2\pi}}\delta\sqrt{\eta}\exp\bigg(-\frac{(u+1)^2\eta}{2}\bigg) + \frac{1}{\sqrt{2\pi}}(1-\delta)\sqrt{\eta}\exp\bigg(-\frac{(u-1)^2\eta}{2}\bigg). \tag{66}$$

Since $S=1, \xi=\eta, \sigma=1, \tilde{\pi}=\pi$, from (35), we obtain

$$\mathcal{G}(-1)\bigg|_{S=1,\sigma=1,\tilde{\pi}=\pi} = \bar{G}(-1,\eta,\alpha), \tag{67}$$

where

$$\begin{aligned}
&\bar{G}(-1,\eta,\alpha)\\
&:= -\int_{-\infty}^{\infty}\bigg((1-\alpha)\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u+1)^2\bigg]\\
&\quad+\alpha\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u-1)^2\bigg]\bigg)\\
&\quad\times\log\bigg((1-\alpha)\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u+1)^2\bigg]\\
&\quad+\alpha\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u-1)^2\bigg]\bigg)du\\
&\quad+\frac{1}{2\beta}\bigg[(\eta-1)\log e-\log\eta\bigg]-\frac{1}{2}\log\frac{2\pi}{\eta}\\
&\quad-\frac{1}{2}\log e+\frac{1}{2\beta}\log(2\pi)+\frac{1}{2\beta}\log e.
\end{aligned} \tag{68}$$

By the symmetry, it is not hard to see that

$$\mathcal{G}(1)\bigg|_{S=1,\sigma=1,\tilde{\pi}=\pi} = \bar{G}(1,\eta,\delta), \tag{69}$$

where

$$\begin{aligned}
&\bar{G}(1,\eta,\delta)\\
&:= -\int_{-\infty}^{\infty}\bigg(\delta\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u+1)^2\bigg]\\
&\quad+(1-\delta)\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u-1)^2\bigg]\bigg)\\
&\quad\times\log\bigg(\delta\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u+1)^2\bigg]\\
&\quad+(1-\delta)\sqrt{\frac{\eta}{2\pi}}\exp\bigg[-\frac{\eta}{2}(u-1)^2\bigg]\bigg)du\\
&\quad+\frac{1}{2\beta}\bigg[(\eta-1)\log e-\log\eta\bigg]-\frac{1}{2}\log\frac{2\pi}{\eta}\\
&\quad-\frac{1}{2}\log e+\frac{1}{2\beta}\log(2\pi)+\frac{1}{2\beta}\log e.
\end{aligned} \tag{70}$$

Now, let $\mathcal{C}_\beta(\alpha,\delta)$ is the set of all solutions η of the equation (64) given β and α and δ. Then, by Claim 1 and (63), the free energy can be expressed as

$$\mathcal{F}\bigg|_{S=1,\sigma=1,\tilde{\pi}=\pi} = \min_{\eta\in\mathcal{C}_\beta(\alpha,\delta)}\bigg[\frac{\delta}{\alpha+\delta}\mathcal{G}(-1)+\frac{\alpha}{\alpha+\delta}\mathcal{G}(1)\bigg], \tag{71}$$

where $\mathcal{G}(-1)$ and $\mathcal{G}(1)$ are given in (67) and (69), respectively.

Solving the optimization problem in (71) is very challenging since $C_\beta(\alpha,\delta)$ may have more than one elements, which corresponds to multiple fixed points of the optimization problem in (71). However, by observing that given α and δ, then β, $\bar{G}(-1,\eta,\alpha)$ and $\bar{G}(1,\eta,\alpha)$ are functions of η. The multiple fixed-points happen if there exists at least two different values η_1 and η_2 such that $\beta(\eta_1)=\beta(\eta_2)$. In simulations, for a fixed β, we can estimate all the values of η such that $|\beta(\eta)-\beta|<10^{-3}$ and then estimate the free energy as a functions of η and find the minimum value among them as the free energy corresponding to $\beta(\eta)$. This procedure can avoid the multiple fixed-point problem.

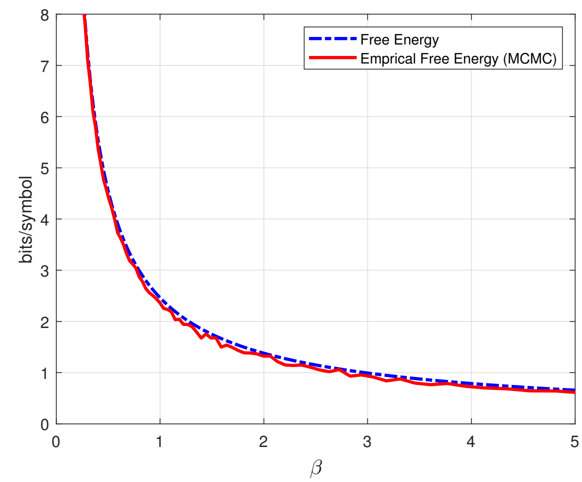

Fig. 3. Free energy by Replica Method and MCMC as functions for the i.i.d. prior $\alpha=\delta=0.5$.

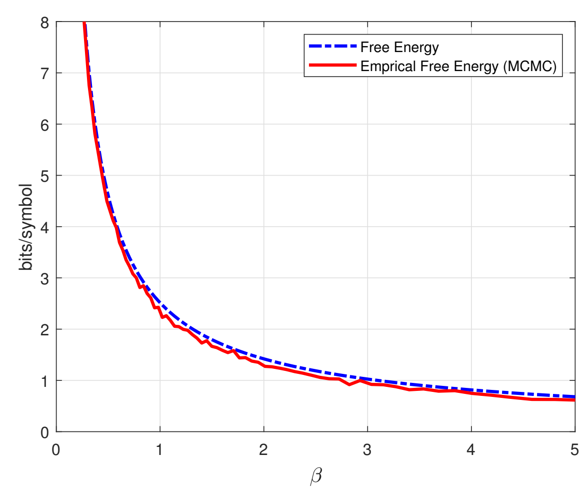

Fig. 4. Free energy by Replica Method and MCMC as functions of β for the symmetric case $\alpha=\delta=0.3$.

2) Markov Chain Monte Carlo (MCMC) vs. Replica Prediction: In this subsection, we use the Markov Chain Monte-Carlo (MCMC) simulation method to estimate the density function $\boldsymbol{P}_{\boldsymbol{y}|\boldsymbol{\Phi}}(\boldsymbol{y}|\boldsymbol{\Phi})$ and verify our replica predictions in Claims 1 and 2. More specifically, we compare the free energies achieved by the replica prediction and MCMC for the linear model with binary-valued Markov prior defined in (62). Our simulation shows that the free energy curves by the replica method and MCMC nearly coincide to each other for all three cases: (1) i.i.d. prior ($\alpha = \delta = 0.5$), (2) symmetric Markov prior $\alpha = \delta = 0.3$, (3) asymmetric Markov prior ($\alpha = 0.2, \delta = 0.5$) (cf. Figs. 3, 4, and 5). In those simulations, the Metropolis–Hastings algorithm is used where the state $\boldsymbol{x}_t := \mathrm{vec}(\Phi_{t+1}, \mathrm{y}_{t+1})$ and the probability transition $g(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t) \sim \mathcal{N}(\boldsymbol{x}_t, \boldsymbol{I}_{mn+n})$. Our simulation results show that the replica prediction for free energy in Claim 1 is very closed to MCMC result.

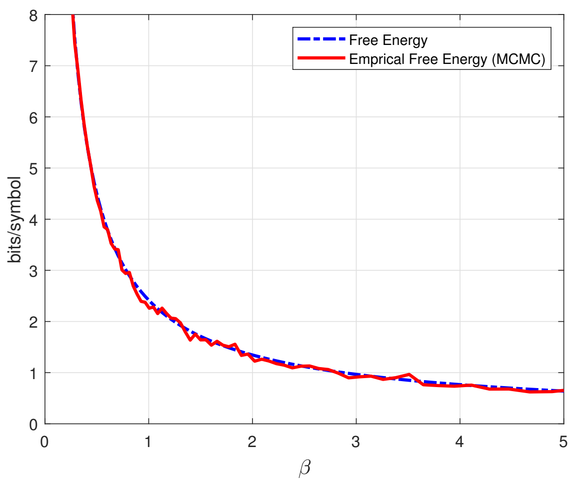

Fig. 5. Free energy by Replica Method and MCMC as functions of β for the non-symmetric case $\alpha = 0.2$ and $\delta = 0.5$.

Since MMSE is fixed function of the free energy (or mutual information) [21], these simulation results also indicate that our replica prediction for MMSE in Claim 2 closely approximates the MMSE of the model.

B. Gauss-Markov Prior

We consider a Gauss-Markov prior $\{X_n\}_{n=1}^\infty$ on $\mathcal{X} = \mathbb{R}$, i.e., $X_n = \nu X_{n-1} + Z_n$, where $Z_n \sim \mathcal{N}(0, \sigma_0^2)$ and $\nu \in (0,1)$. Then, the transition probability is

$$\pi(x_0, x) := \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_0^2}(x - \nu x_0)^2\right]. \tag{72}$$

This means that $X_n|X_{n-1} = x_0 \sim \mathcal{N}(\nu x_0, \sigma_0^2)$ for all $n \in \mathbb{Z}_+$. This is not hard to show that the Markov chain in (72) is irreducible by using [42, Definition 1.1]. We even can show that this Markov chain is a Harris chain by using its definition in [46] or using [42, Theorem 4.2]. To guarantee the irreducible and recurrent properties of this continuous-space Markov chain, we show that $\mathbb{P}[\tau_A < \infty | X_0 = x] = 1$ for any $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, where $\tau_A = \{\inf n \geq 1 : X_n \in \mathcal{A}\}$.

1) Free Energy and Average Mutual Information: We assume that all postulated distributions are the same as their true ones for simplicity. Now, given $S = s_0 \in \mathbb{R}_+$[6], by using Claim 1, we can show that the free energy satisfies

$$\begin{aligned}\mathcal{F}_q\Big|_{S=s_0,\sigma=1,\tilde{\pi}=\pi} &= \frac{1}{2}\log\left(2\pi e\left(s_0\sigma_0^2 + \frac{1}{\eta}\right)\right) + \frac{1}{2\beta}\left[(\eta - 1)\log e - \log \eta\right] \\ &\quad - \frac{1}{2}\log\frac{2\pi}{\eta} - \frac{1}{2}\log e + \frac{1}{2\beta}\log(2\pi) + \frac{1}{2\beta}\log e \end{aligned}\tag{73}$$

where

$$\eta = \frac{-((\beta-1)s_0\sigma_0^2 + 1) + \sqrt{((\beta-1)s_0\sigma_0^2+1)^2 + 4s_0\sigma_0^2}}{2s_0\sigma_0^2}. \tag{74}$$

2) Markov Chain Monte Carlo (MCMC) vs. Replica Prediction: In this subsection, we use the same MCMC algorithm as Subsection IV-A2, which is the Metropolis–Hastings algorithm. In the Fig. 6, we plot the free energy curves for the linear model with Markov prior in (72) for three cases $\nu = 0.1$, $\nu = 0.5$, and $\nu = 0.8$. The curves suggest that the free energy does not depend on ν as we can observe from (73). In these plots, we set $X_1 \sim \mathcal{N}(0, \frac{\sigma_0^2}{1-\nu^2})$ to force the state distribution of the Markov (Harris) chain $X_n \sim \mathcal{N}(0, \frac{\sigma_0^2}{1-\nu^2})$ for all $n \geq 1$. The plot also shows that the replica prediction for the free energy is very closed to the MCMC simulation result. Since the MMSE is a fixed function of the free energy (or mutual information) [21], this also means that the MMSE curve by replica method closely approaches the MMSE of the model.

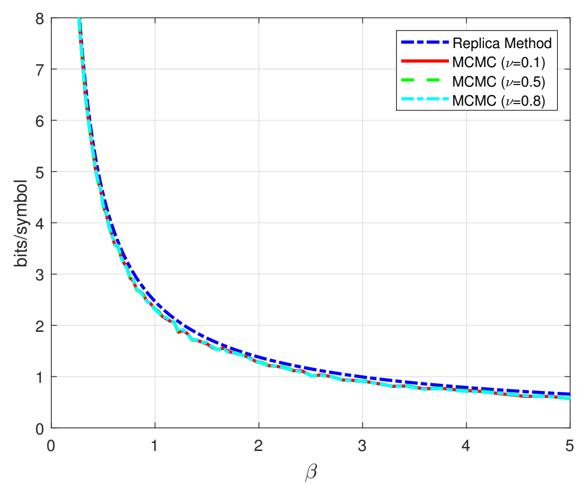

Fig. 6. Free energy by replica method and empirical MCMC as functions of β for $\sigma_0^2 = 1$ and $s_0 = 1$.

C. Hidden Markov Prior

In this section, we estimate free energy and mutual information for the linear model in Section II with hidden Markov sources defined in [32, Sect. 7]. The sequence $\{X_n\}_{n=1}^\infty$ which takes values on $\mathbb{R}$ is generated via

$$\begin{aligned} &p_{X_n|\Upsilon_n}(x_n \mid \upsilon_n) \\ &\quad = \upsilon_n \mathcal{N}(x_n; 0, 1) + (1 - \upsilon_n)\delta(x_n) \quad (75) \\ &\quad = \frac{\upsilon_n}{\sqrt{2\pi}}\exp\left(-\frac{x_n^2}{2}\right) + (1-\upsilon_n)\delta(x_n) \quad (76) \end{aligned}$$

[6]For example, in BPSK or QPSK modulation schemes in communications, all symbols in the constellation have a fixed energy s_0.

using a time-homogeneous irreducible Markov chain-generated sparsity pattern $\{\Upsilon_n\}_{n=1}^{\infty}$ on $\mathcal{S}_\gamma = \{0,1\}$. Such a Markov chain is fully described by the following transition stochastic matrix

$$P_\Upsilon = \begin{bmatrix} 1-\kappa\gamma & \gamma\kappa \\ (1-\kappa)\gamma & 1-(1-\kappa)\gamma \end{bmatrix} \tag{77}$$

for some $\gamma \in (0,1]$ called the Markov independence parameter. This irreducible Markov chain yields a stationary distribution with activity rate $P(\Upsilon_n = 1) = \kappa$ for all $n \in \mathbb{Z}^+$.

1) Free Energy and Average Mutual Information: First, it is easy to see that the left Perron-Frobenius eigenvector of the stochastic matrix P_Υ with unit Manhattan norm is

$$\lambda_0 = (1-\kappa, \kappa)^T. \tag{78}$$

Observe that

$$\begin{aligned} &P_{\mathsf{X}_1,\Upsilon_1|\mathsf{X}_0,\Upsilon_0}(x_1, \upsilon_1 \mid x_0, \upsilon_0) \\ &\quad = P_{X|\Upsilon}(x_1 \mid \upsilon_1)\pi_\Upsilon(\upsilon_0, \upsilon_1) \end{aligned} \tag{79}$$

where $P_{X|\Upsilon}(\cdot|\cdot)$ is the emission probability of the hidden Markov process.

For this case, $\eta = \xi$, and (56)–(57) is degraded to the following equation in η:

$$\eta^{-1} = 1 + \beta(1-\kappa)g(\gamma k) + \beta\kappa g(1-(1-\kappa)\gamma), \tag{80}$$

where

$$g(x) := x - \mathbb{E}\left[\left(\frac{L_1(x,U)}{L_2(x,U)}\right)^2\right], \tag{81}$$

and

$$L_1(x,U) := x\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left(-\frac{\eta U^2}{2(1+\eta)}\right)\left(\frac{\eta U}{1+\eta}\right), \tag{82}$$

$$\begin{aligned} L_2(x,U) := &(1-x)\sqrt{\frac{\eta}{2\pi}}\exp\left(-\frac{\eta U^2}{2}\right) \\ &+ x\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left(-\frac{\eta U^2}{2(1+\eta)}\right). \end{aligned} \tag{83}$$

Now, let $\hat{C}_\beta(\kappa,\gamma)$ is the set of all solutions η of equation (80) given β and κ and γ. By using Claim 3, it can be shown that

$$\begin{aligned} &\mathcal{F}\Big|_{S=1,\sigma=1,\tilde{\pi}_\Upsilon=\pi_\Upsilon} \\ &= \min_{\eta \in \hat{C}_\beta(\kappa,\gamma)} \left[(1-\kappa)\hat{G}(0,\eta,\kappa,\gamma) + \kappa\hat{G}(1,\eta,\kappa,\gamma)\right], \end{aligned} \tag{84}$$

where

$$\begin{aligned} &\hat{G}(0,\eta,\kappa,\gamma) \\ &:= -\int_{-\infty}^{\infty}\Bigg((1-\kappa\gamma)\sqrt{\frac{\eta}{2\pi}}\exp\left[-\frac{\eta u^2}{2}\right] \\ &\quad + \kappa\gamma\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left[-\frac{\eta u^2}{2(1+\eta)}\right]\Bigg) \\ &\quad \times \log\Bigg((1-\kappa\gamma)\sqrt{\frac{\eta}{2\pi}}\exp\left[-\frac{\eta u^2}{2}\right] \\ &\quad + \kappa\gamma\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left[-\frac{\eta u^2}{2(1+\eta)}\right]\Bigg)du \\ &\quad + \frac{1}{2\beta}\Big[(\eta-1)\log e - \log\eta\Big] - \frac{1}{2}\log\frac{2\pi}{\eta} \\ &\quad - \frac{1}{2}\log e + \frac{1}{2\beta}\log(2\pi) + \frac{1}{2\beta}\log e. \end{aligned} \tag{85}$$

and

$$\begin{aligned} &\hat{G}(1,\eta,\kappa,\gamma) \\ &:= -\int_{-\infty}^{\infty}\Bigg((1-\kappa)\gamma\sqrt{\frac{\eta}{2\pi}}\exp\left[-\frac{\eta u^2}{2}\right] \\ &\quad + (1-(1-\kappa)\gamma)\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left[-\frac{\eta u^2}{2(1+\eta)}\right]\Bigg) \\ &\quad \times \log\Bigg((1-\kappa)\gamma\sqrt{\frac{\eta}{2\pi}}\exp\left(-\frac{\eta u^2}{2}\right) \\ &\quad + (1-(1-\kappa)\gamma)\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left[-\frac{\eta u^2}{2(1+\eta)}\right]\Bigg)du \\ &\quad + \frac{1}{2\beta}\Big[(\eta-1)\log e - \log\eta\Big] - \frac{1}{2}\log\frac{2\pi}{\eta} \\ &\quad - \frac{1}{2}\log e + \frac{1}{2\beta}\log(2\pi) + \frac{1}{2\beta}\log e. \end{aligned} \tag{86}$$

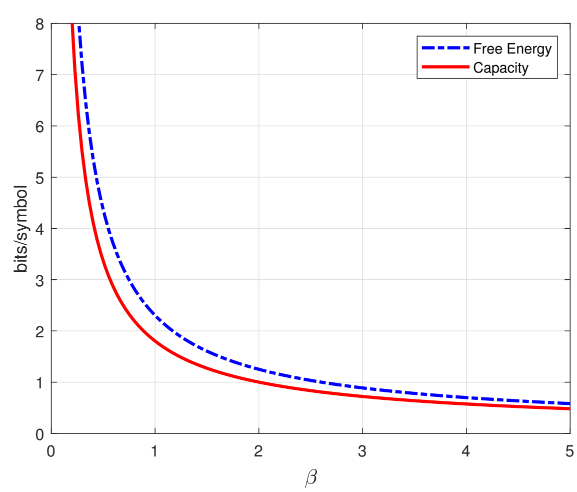

Fig. 7. Free energy and average mutual information as functions of β for the symmetric i.i.d. case $\lambda = 0.5$ and $\gamma = 1$.

Solving the optimization problems in (84) is very challenging. However, by observing that given κ and γ, then β, $\hat{G}(0,\eta,\kappa,\gamma)$ and $\hat{G}(1,\eta,\kappa,\gamma)$ are functions of η. Hence, we can plot lower and upper bounds for the free energy $\mathcal{F}$ and the average mutual information as functions of (κ,γ). In Fig. 7, we plot the free energy and the average mutual information for $\kappa=0.5$ and $\gamma=1$, i.e., the sequence $\{X_n\}_{n=1}^{\infty}$ is i.i.d. generated.

For the non-symmetric case where $\kappa=0.3$ and $\gamma=0.8$, we obtain the free energy and the average mutual information as in Fig. 8.

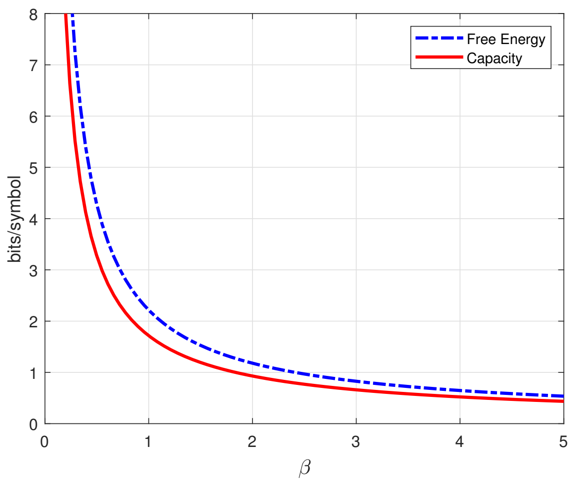

Fig. 8. Free energy and average mutual information as functions of β for the non-symmetric case $\lambda=0.3$ and $\gamma=0.8$.

2) Approximate Message Passing Algorithm vs. Replica Prediction: Denote by

$$K_1(x,U):=x\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left(-\frac{\eta U^2}{2(1+\eta)}\right)\left(\frac{\eta U}{1+\eta}\right), \tag{87}$$

$$\begin{aligned}K_2(x,U):=&(1-x)\sqrt{\frac{\eta}{2\pi}}\exp\left(-\frac{\eta U^2}{2}\right)\\&+x\sqrt{\frac{\eta}{2\pi(1+\eta)}}\exp\left(-\frac{\eta U^2}{2(1+\eta)}\right),\quad\forall x,U.\end{aligned} \tag{88}$$

Let

$$R_1:=\mathbb{E}\left[\left(\frac{K_1(\kappa\gamma,U)}{K_2(\kappa\gamma,U)}\right)^2\right] \tag{89}$$

and

$$R_2:=\mathbb{E}\left[\left(\frac{K_1(1-(1-\kappa)\gamma,U)}{K_2(1-(1-\kappa)\gamma,U)}\right)^2\right]. \tag{90}$$

Then, by using Claim 3, we can show that

$$\mathrm{MMSE}_{\mathrm{HM}}=\kappa-((1-\kappa)R_1+\kappa R_2). \tag{91}$$

In this section, we compare the MMSE in Claim 3 with the MSE achieved by the AMP algorithm in [32] for $n=1000$ (signal dimension) and $m=\lceil\frac{n}{\beta}\rceil$ (observations). We assume that $S=1$ and $\boldsymbol{A}$ is a random matrix where each element is normal distributed $\mathcal{N}(0,1/m)$ as Section II. However, this algorithm assumes some level of sparsity in signal $\boldsymbol{X}$. Before introducing the algorithm, we define some new functions:

$$\alpha_l(c)=\frac{1}{c+1}, \tag{92}$$

$$\beta_l(c)=\left(\frac{1-\kappa}{\kappa}\right)\left(\frac{c+1}{c}\right), \tag{93}$$

$$\zeta_l(c)=\frac{1}{c(c+1)}, \tag{94}$$

$$F_l(\theta;c)=\frac{\alpha_l(c)\theta}{1+\beta_n(c)e^{-\zeta_l(c)|\theta|^2}}, \tag{95}$$

$$G_l(\theta;c)=\beta_n(c)e^{-\zeta_n(c)|\theta|^2}|F_n(\theta;c)|^2+\frac{c}{\theta}F_l(\theta;c) \tag{96}$$

$$\begin{aligned}F_l'(\theta;c)=&\frac{\alpha_l(c)}{1+\beta_l(c)e^{-\zeta_l(c)|\theta|^2}}\\&\times\left[1+\frac{\zeta_l(c)|\theta^2|}{1+(\beta_l(c)e^{-\zeta_l(c)|\theta|^2})^{-1}}\right],\quad\forall l\in[n].\end{aligned} \tag{97}$$

We call this algorithm Turbo AMP since it is based on an approximation of a loopy BP which has demonstrated very accurate results in LDPC and Turbo decoding [32]. The algorithm for our setting is as follows:

1) Initialize

$$c^0=10;\quad \mu_l^0=0\quad\forall l\in[n];\quad z_k^0=y_k\ \ \forall k\in[m]. \tag{98}$$

2) Repeat the following for all $i=0,1,2,\cdots$ (we use 10 iterations in our simulations):

$$\theta_l^i=\frac{1}{\sqrt{n}}\sum_{k=1}^m A_{kl}z_k^i+\mu_l^i,\quad\forall l\in[n], \tag{99}$$

$$\mu_l^{i+1}=F_l(\theta_l^i;c^i),\quad\forall l\in[n], \tag{100}$$

$$v_l^{i+1}=G_l(\theta_l^i;c^i),\quad\forall l\in[n], \tag{101}$$

$$c^{i+1}=1+\frac{\beta}{n}\sum_{l=1}^n v_l^{i+1}, \tag{102}$$

$$\begin{aligned}z_k^{i+1}=&y_k-\sum_{l=1}^n A_{kl}\mu_l^{i+1}+\frac{z_k^i}{m}\sum_{l=1}^n F_l'(\theta_l^i;c^i),\\&\forall k\in[m].\end{aligned} \tag{103}$$

Our obtained results are as follows.

- For the symmetric case $\kappa=0.5$ and $\gamma=1$, the Markov model in Section II becomes the linear model with i.i.d. sequence $\{X_n\}_{n=1}^{\infty}$ in [4, Sect. II]. Fig. 9 shows that Turbo AMP works well for this case. The gap between the MSE of AMP and the MSE of the Replica Method in Claim 3 is very small.
- For the non-symmetric case $\kappa=0.3$ and $\gamma=0.8$, the Markov model in Section II is very different from the linear model with i.i.d. sequence $\{X_n\}_{n=1}^{\infty}$ in [4, Sect. II]. Fig. 10 shows that Turbo AMP also works well for this case. The gap between the MSE of Turbo AMP and the upper bound of MSE by using the Replica Method in Claim 3 is also still small. However, the gap is bigger than the symmetric case. The multiple fixed points (multiple solutions) of the equation (80) can be a reason for this

gap. Besides, Turbo AMP may not be optimal for this given model although it exploits the Markov structure of the sequence $\{X_n\}_{n=1}^{\infty}$ quite well.

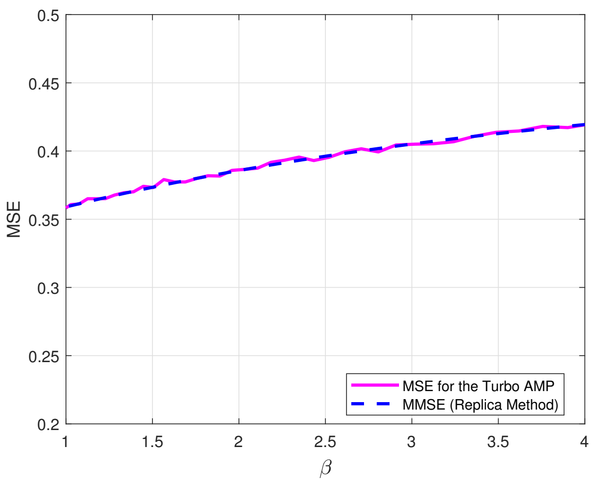

Fig. 9. MSE by Turbo AMP algorithm and MMSE by the Replica Method as functions of β for the symmetric case $\kappa = 0.5, \gamma = 1$, i.e., $\{X_n\}_{n=1}^{\infty}$ is an i.i.d. sequence.

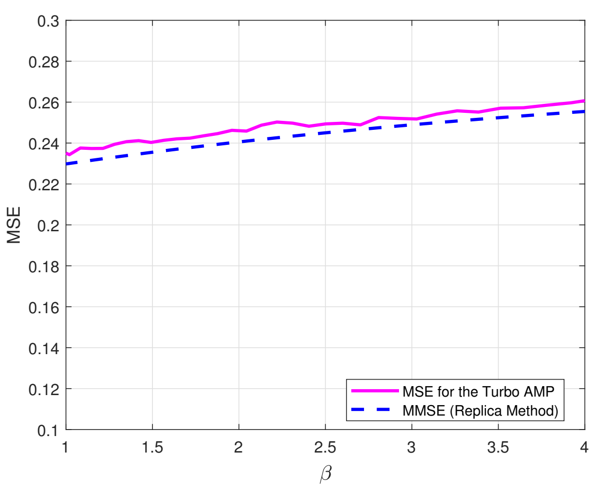

Fig. 10. MSE by Turbo AMP algorithm and MMSE by the Replica Method as functions of β for the asymmetric case $\kappa = 0.3$ and $\gamma = 0.8$.

V. Proofs of Main Results

This section proves Claims 1–3 using the replica method. We first prove some preliminary results which are required to estimate the free energy of the linear model with Markov signal prior. Then, we derive the joint moments for this model. Finally, we obtain the free energy and joint moments for the linear model with hidden Markov signal prior based on the results of the linear model with Markov signal prior.

A. Some preliminary results

Lemma 4. *[4, p. 1998] Let $X_n^{(a)}$ be replicated vectors with distribution $q_{\boldsymbol{X}}$. Define a sequence of $(\nu+1)\times(\nu+1)$ random matrices $\{\boldsymbol{Q}_n\}_{n=1}^{\infty}$ such that*

$$Q_n^{(a,b)} = S_n X_n^{(a)} X_n^{(b)} \tag{104}$$

for all $a, b \in \{0, 1, \cdots, \nu\}$ and $n = 1, 2, \cdots$. Let

$$\boldsymbol{T}_n = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{Q}_k, \quad n = 1, 2, \cdots. \tag{105}$$

Then, the following holds:

$$\begin{aligned} &\frac{1}{n} \log \mathbb{E}[Z^{\nu}(\boldsymbol{Y}, \boldsymbol{\Phi})] \\ &\quad = \frac{1}{n} \log \mathbb{E}\bigg\{ \exp\Big[m\Big(G^{(\nu)}(\boldsymbol{T}_n) + O(n^{-1}) \Big) \Big] \bigg\}, \end{aligned} \tag{106}$$

where

$$\begin{aligned} G^{(\nu)}(Q) &:= -\frac{1}{2} \log \det(I + \Sigma Q) \\ &\quad - \frac{1}{2} \log\left(1 + \frac{\nu}{\sigma^2}\right) - \frac{\nu}{2} \log(2\pi\sigma^2), \end{aligned} \tag{107}$$

and Σ is a $(\nu+1)\times(\nu+1)$ matrix

$$\Sigma = \frac{\beta}{\sigma^2 + \nu} \begin{bmatrix} \nu & -e^T \\ -e & (1 + \frac{\nu}{\sigma^2})I - \frac{1}{\sigma^2} e e^T \end{bmatrix}, \tag{108}$$

where e is a $\nu \times 1$ column vector whose entries are all 1.

The following two lemmas state some new results on large deviations for Markov chains induced by the channel setting.

Lemma 5. *Let $\{S_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of random variable on a finite set $\mathcal{S} \subset \mathbb{R}^+$. Let $\boldsymbol{X} := \{X_n\}_{n=1}^{\infty}$ be a Markov chain with states on a Polish space $\mathcal{X}$ with the transition matrix $P = \{\pi(x, x')\}_{x, x' \in \mathcal{X}}$. Assume this Markov chain is irreducible. Set $\boldsymbol{X}^{(0)} = \boldsymbol{X}$. Let $\boldsymbol{X}^{(a)} := \{X_n^{(a)}\}_{n=1}^{\infty}$ be a set of ν replica sequences with (postulated) distribution $q_{\boldsymbol{X}}$ for each $a = 1, 2, \cdots, \nu$. This means that*

$$\begin{aligned} &p_{\boldsymbol{X}^{(0)}\boldsymbol{X}^{(1)}\boldsymbol{X}^{(2)}\cdots\boldsymbol{X}^{(\nu)}}(x^{(0)}, x^{(1)}, x^{(2)}, \cdots, x^{(\nu)}) \\ &\quad \sim p_{\boldsymbol{X}}(x^{(0)}) \prod_{i=1}^{\nu} q_{\boldsymbol{X}}(x^{(i)}), \end{aligned} \tag{109}$$

where

$$p_{\boldsymbol{X}}(x^{(0)}) = \prod_{i=1}^{\infty} \pi(x_i^{(0)}, x_{i+1}^{(0)}) \tag{110}$$

$$q_{\boldsymbol{X}}(x^{(a)}) = \prod_{i=1}^{\infty} \tilde{\pi}(x_i^{(a)}, x_{i+1}^{(a)}), \qquad \forall a \in [\nu]. \tag{111}$$

Define a new sequence of $(\nu+1)\times(\nu+1)$ random matrices $\{\boldsymbol{Q}_n\}_{n=1}^{\infty}$ such that

$$Q_n^{(a,b)} = S_n X_n^{(a)} X_n^{(b)} \tag{112}$$

for all $a \in [\nu]$ and $b \in [\nu]$ and for all $n = 1, 2, \cdots$. Then, $\{\boldsymbol{Q}_n\}_{n=1}^{\infty}$ is also an irreducible Markov chain with states on $\mathcal{Q}$, where $\mathcal{Q}$ is defined in (3). In addition, the transition probability, namely $P(Q|Q')$, of this Markov chain satisfies (113), where $p_{X_{n-1}}(\cdot)$ is the state distribution at time $n-1$ of the Markov chain $\{X_n\}_{n=1}^{\infty}$ with the transition probability π defined in (6), and $q_{X_{n-1}}(\cdot)$ is the state distribution at time

$$P(Q|Q') = \frac{\sum_{(s,x_0,x_1,\cdots,x_\nu,s',x_0',x_1',\cdots,x_\nu')\in\mathcal{A}_Q\times\mathcal{A}_{Q'}} P_S(s')P_S(s)p_{X_{n-1}}(x_0')\pi(x_0',x_0)\prod_{i=1}^{\nu} q_{X_{n-1}}(x_i')\tilde{\pi}(x_i',x_i)}{\sum_{(s',x_0',x_1',\cdots,x_\nu')\in\mathcal{A}_{Q'}} P_S(s')p_{X_{n-1}}(x_0')\prod_{i=1}^{\nu} q_{X_{n-1}}(x_i')} \tag{113}$$

$n-1$ of the Markov chain $\{X_n\}_{n=1}^\infty$ with the (postulated) transition probability $\tilde{\pi}(\cdot,\cdot)$ defined in (11), and

$$\mathcal{A}_Q := \{(s,x)\in\mathcal{S}\times\mathcal{X}^{\nu+1} : sxx^T = Q\}, \quad \forall Q\in\mathcal{Q}. \tag{114}$$

The proof of Lemma 5 is based on showing $\mathbb{P}(\boldsymbol{Q}_n = Q|\sigma(\boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_{n-1})) = \mathbb{P}(\boldsymbol{Q}_n = Q|\boldsymbol{Q}_{n-1})$ by using $\sigma(\boldsymbol{Q}_k) = \sigma(S_k, X_k^{(0)}, X_k^{(1)},\cdots,X_k^{(\nu)})$ for all $k\in[n]$[7].

Lemma 6. *Let $\mathcal{X}$ be a Polish space with finite cardinality and a irreducible Markov chain $\boldsymbol{X} := \{X_n\}_{n=1}^\infty$ defined on $\mathcal{X}$ and ν be a positive integer number. Let $X_n^{(a)}$ for $a\in[\nu]$ be replicas of the Markov process $\boldsymbol{X}$. Recall the definition of the sequence $\boldsymbol{Q}_n$ in Lemma 5 and $\boldsymbol{T}_n = \frac{1}{n}\sum_{j=1}^n \boldsymbol{Q}_j$. Let $P_n(U) := \mathbb{P}(\boldsymbol{T}_n\in U)$ for any measurable set U on the σ-algebra generated by $\{\boldsymbol{Q}_n\}_{n=1}^\infty$. Then, for and bounded and continuous function $F:\mathcal{Q}\to\mathbb{R}$*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\big[e^{nF(\boldsymbol{T}_n)}\big]$$

$$= \lim_{n\to\infty}\frac{1}{n}\log\int e^{nF(Q)}dP_n(Q) \tag{115}$$

$$= \sup_Q\Big[F(Q) - I(Q)\Big] \tag{116}$$

where $I(Q) = \sup_{\tilde{Q}}(\mathrm{tr}(\tilde{Q}Q) - \log\rho(P_{\tilde{Q}}))$ and $\rho(P_{\tilde{Q}})$ is the Perron-Frobenius eigenvalue of the matrix $P_{\tilde{Q}} = \{e^{\mathrm{tr}(\tilde{Q}\bar{Q}_j)}P_{\bar{Q}_j|\bar{Q}_i}\}_{0\le i,j\le M}$ and $M = |\mathcal{Q}|-1$, where $\mathcal{Q}$ and $\{\bar{Q}_i\}_{i=0}^M$ are defined in Subsection I-D.

Proof: Equation (115) is an application of the change of measures [48]. Equation (116) is a direct application of the large deviation theorem [49]. ∎

Lemma 7. *Recall the definitions of $\{\bar{Q}_i\}_{i=0}^M$ in Subsection I-D Recall the definition of the sequence $\boldsymbol{Q}_n$ in Lemma 5. Then, the following holds:*

$$\frac{\partial\log\rho(P_{\tilde{Q}})}{\partial\tilde{Q}}(\tilde{Q})$$

$$= \frac{1}{\rho(P_{\tilde{Q}})}\sum_{i=0}^M\lambda_i(\tilde{Q})\sum_{j=0}^M\psi_j(\tilde{Q})\bar{Q}_jP(\bar{Q}_j|\bar{Q}_i)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_j)}, \tag{117}$$

where $\lambda(\tilde{Q})$ and $\psi(\tilde{Q})$ are left and right eigenvectors associated with the Perron-Frobenius eigenvalue $\rho(P_{\tilde{Q}})$ which are normalized such that $\lambda(\tilde{Q})^T\psi(\tilde{Q}) = 1$.

Proof: Refer to Appendix A for a detailed proof. ∎

[7]A detailed proof for this Lemma can be found in [47, Proof of Lemma 29].

Theorem 8. *Recall the definition of $G^{(\nu)}(Q)$ in Lemma 4. In the large system limit, given any initial state x_0, the free energy satisfies:*

$$\mathcal{F}_q\big|_{X_0=x_0} = -\lim_{\nu\to0}\frac{\partial}{\partial\nu}\sup_Q\Big[\beta^{-1}G^{(\nu)}(Q) - I^{(\nu)}(Q)\Big], \tag{118}$$

where

$$I^{(\nu)}(Q) := \sup_{\tilde{Q}}\Big[\mathrm{tr}(\tilde{Q}Q) - \log\rho(P_{\tilde{Q}})\Big], \tag{119}$$

and $\rho(P_{\tilde{Q}})$ is the Perron-Frobenius eigenvalue of the matrix $P_{\tilde{Q}} = \{e^{\mathrm{tr}(\tilde{Q}\bar{Q}_j)}P_{\bar{Q}_j|\bar{Q}_i}\}_{0\le i,j\le M}$ and $M = |\mathcal{Q}|-1$.

Proof: The proof follows the same idea as [4, Part A, Sect. IV] with some important changes to account for the Markov setting.

1) By applying Lemma 6 and [4, p. 1998], we obtain

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[Z^\nu(\boldsymbol{Y},\boldsymbol{\Phi})]$$

$$= \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\bigg\{\exp\bigg[\frac{n}{\beta}\Big(G^{(\nu)}(\boldsymbol{T}_n) + O(n^{-1})\Big)\bigg]\bigg\} \tag{120}$$

$$= \sup_Q\Big[\frac{1}{\beta}G^{(\nu)}(Q) - I^{(\nu)}(Q)\Big]. \tag{121}$$

2) Estimate the free energy.

Now, observe that

$$\mathcal{F}_q\big|_{X_0=x_0}$$

$$= -\lim_{n\to\infty}\frac{1}{n}\lim_{\nu\to0}\frac{\partial}{\partial\nu}\log\mathbb{E}[Z^\nu(\boldsymbol{Y},\boldsymbol{\Phi})] \tag{122}$$

$$= -\lim_{\nu\to0}\frac{\partial}{\partial\nu}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[Z^\nu(\boldsymbol{Y},\boldsymbol{\Phi})] \tag{123}$$

$$= -\lim_{\nu\to0}\frac{\partial}{\partial\nu}\sup_Q\Big[\frac{1}{\beta}G^{(\nu)}(Q) - I^{(\nu)}(Q)\Big]\bigg|_{X_0=x_0} \tag{124}$$

where (122) follows from the assumption (A1), (A2), and the fact (F1), (123) follows from the assumption (A3), and (124) follows from (121).

∎

Theorem 9. *Recall the definitions of Σ in Lemma 4, the matrix $P_{\tilde{Q}}$ in Theorem 8, and $\{\bar{Q}_i\}_{i=0}^M$ in Subsection I-D. The optimal* *matrix Q^* of equation (118) in Theorem 8 must satisfy the following constraints:*

$$Q^* = \frac{\partial \log \rho(P_{\tilde{Q}^*})}{\partial \tilde{Q}^*}, \tag{125}$$
$$\tilde{Q}^* = -(2\beta)^{-1}(I+\Sigma Q^*)^{-1}\Sigma, \tag{126}$$
$$\frac{\partial \log \rho(P_{\tilde{Q}^*})}{\partial \tilde{Q}^*} = \frac{1}{\rho(P_{\tilde{Q}^*})}\sum_{i=0}^{M}\lambda_i(\tilde{Q}^*)\sum_{j=0}^{M}\psi_j(\tilde{Q}^*)\bar{Q}_j P(\bar{Q}_j|\bar{Q}_i)e^{\mathrm{tr}(\tilde{Q}^*\bar{Q}_j)}, \tag{127}$$

where $\lambda(\tilde{Q}^*)$ and $\psi(\tilde{Q}^*)$ are left and right eigenvectors associated with the Perron-Frobenius eigenvalue $\rho(P_{\tilde{Q}^*})$ which are normalized such that $\lambda(\tilde{Q}^*)^T\psi(\tilde{Q}^*) = 1$.

Proof: Recall the definition of $I^{(\nu)}$ in Theorem 8. It is easy to see that the optimization problem in (121) is equivalent to the following optimization problem:

$$\sup_{Q}\inf_{\tilde{Q}} T^{(\nu)}(Q,\tilde{Q}) \tag{128}$$

where

$$T^{(\nu)}(Q,\tilde{Q}) := -\frac{1}{2\beta}\log\det(I+\Sigma Q) - \mathrm{tr}(\tilde{Q}Q) + \log\rho(P_{\tilde{Q}}) - \frac{1}{2\beta}\log\left(1+\frac{\nu}{\sigma^2}\right) - \frac{\nu}{2\beta}\log(2\pi\sigma^2). \tag{129}$$

For an arbitrary Q, we first seek critical points with respect to $\tilde{Q}$ and find that for any given Q, the extremum in $\tilde{Q}$ satisfies

$$Q = \frac{\partial \log \rho(P_{\tilde{Q}})}{\partial \tilde{Q}} \tag{130}$$

Let $\tilde{Q}(Q)$ be a solution to (130). We then seek the critical point of $T^{(\nu)}(Q,\tilde{Q}(Q))$ with respect to Q.

Let

$$K_{Q,\tilde{Q}} := \begin{bmatrix} \frac{\partial \tilde{Q}_{0,0}}{\partial Q_{0,0}} & \frac{\partial \tilde{Q}_{0,1}}{\partial Q_{0,1}} & \cdots & \frac{\partial \tilde{Q}_{0,\nu}}{\partial Q_{0,\nu}} \\ \frac{\partial \tilde{Q}_{1,0}}{\partial Q_{1,0}} & \frac{\partial \tilde{Q}_{1,1}}{\partial Q_{1,1}} & \cdots & \frac{\partial \tilde{Q}_{1,\nu}}{\partial Q_{1,\nu}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \tilde{Q}_{\nu,0}}{\partial Q_{\nu,0}} & \frac{\partial \tilde{Q}_{\nu,1}}{\partial Q_{\nu,1}} & \cdots & \frac{\partial \tilde{Q}_{\nu,\nu}}{\partial Q_{\nu,\nu}} \end{bmatrix}. \tag{131}$$

Observe that

$$\frac{\partial \mathrm{tr}(\tilde{Q}Q)}{\partial Q} = \frac{\partial \mathrm{tr}(Q\tilde{Q})}{\partial Q} \tag{132}$$
$$= \tilde{Q} + Q \odot K_{Q,\tilde{Q}}, \tag{133}$$

where $\odot$ is the Hadamard product.

It follows that

$$\frac{\partial T^{(\nu)}(Q,\tilde{Q})}{\partial Q} = -\frac{1}{2\beta}(I+\Sigma Q)^{-1}\Sigma - \left(\tilde{Q} + Q\odot K_{Q,\tilde{Q}}\right) + \frac{\partial \log\rho(P_{\tilde{Q}})}{\partial Q} \tag{134}$$
$$= -\frac{1}{2\beta}(I+\Sigma Q)^{-1}\Sigma - \left(\tilde{Q} + Q\odot K_{Q,\tilde{Q}}\right) + \frac{\partial \log\rho(P_{\tilde{Q}})}{\partial \tilde{Q}} \odot K_{Q,\tilde{Q}} \tag{135}$$
$$= -\frac{1}{2\beta}(I+\Sigma Q)^{-1}\Sigma - \tilde{Q} - \left[Q - \frac{\partial \log\rho(P_{\tilde{Q}})}{\partial \tilde{Q}}\right]\odot K_{Q,\tilde{Q}} \tag{136}$$
$$= -\frac{1}{2\beta}(I+\Sigma Q)^{-1}\Sigma - \tilde{Q}, \tag{137}$$

where (134) follows from (133), and (137) follows from (130). Hence, the optimal value of the Theorem 8 is the solution of the following equation systems:

$$Q = \frac{\partial \log \rho(P_{\tilde{Q}})}{\partial \tilde{Q}}, \tag{138}$$
$$\tilde{Q} = -(2\beta)^{-1}(I+\Sigma Q)^{-1}\Sigma. \tag{139}$$

Finally, from Lemma 7, we also obtain an additional constraint in (127). ■

Observe that the matrix Σ defined in Lemma 4 is invariant if two non-zero indices are interchanged, i.e., Σ is symmetric in replicas. Now, we use the RS assumption (A4) to simplify the result in Theorem 8. More specifically, we use the following RS assumption:

Definition 10. *[4, p. 1999] An solution $(\tilde{Q}^*, Q^*)$ of the optimization problem in Theorem 8, i.e.,*

$$\sup_{Q}\left[\beta^{-1}G^{(\nu)}(Q) - I^{(\nu)}(Q)\right] = \sup_{Q}\inf_{\tilde{Q}}\left[-\frac{1}{2\beta}\log\det(I+\Sigma Q) - \mathrm{tr}(\tilde{Q}Q) + \log\rho(P_{\tilde{Q}}) - \frac{1}{2\beta}\log\left(1+\frac{\nu}{\sigma^2}\right) - \frac{\nu}{2\beta}\log(2\pi\sigma^2)\right], \tag{140}$$

is called to satisfy the Replica Symmetry (RS) if both Q^* and $\tilde{Q}^*$ under the exchange of any two (nonzero) replica indices. In other words, the extrema can be written as

$$Q^* = \begin{bmatrix} r & m & m & \cdots & m \\ m & p & q & \cdots & q \\ m & q & p & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & q \\ m & q & \cdots & q & p \end{bmatrix}, \tag{141}$$

$$\tilde{Q}^* = \begin{bmatrix} c & d & d & \cdots & d \\ d & g & f & \cdots & f \\ d & f & g & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & f \\ d & f & \cdots & f & g \end{bmatrix}, \tag{142}$$

where r, m, p, q, c, d, f, g are some real numbers which are not dependent on ν.

Next, we show the following results:

Lemma 11. *Let $\{\bar{Q}_i\}_{i=0}^M$ be states of the Markov chain $\{\boldsymbol{Q}_n\}_{n=1}^\infty$ in Lemma 5. Assume that*

$$\rho(P_{\tilde{Q}^*}) \to 1 \quad \text{and} \quad \sum_{j=0}^M P(\bar{Q}_j|\bar{Q}_i) e^{\operatorname{tr}(\tilde{Q}^*\bar{Q}_j)} \to 1 \tag{143}$$

for all $i \in [M]$ as $\nu \to 0$. Then, under the RS assumption in Definition 10, the following holds:

$$Q^* = \lim_{\nu\to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}^*) \mathbb{E}[\boldsymbol{Q}_1 e^{\operatorname{tr}(\tilde{Q}^*\boldsymbol{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i] \tag{144}$$

where Q^* is defined in Theorem 9 and $\lambda(\tilde{Q}^*)$ is a left (positive) eigenvector associated with the Perron-Frobenius eigenvalue $\rho(P_{\tilde{Q}^*})$ such that $\|\lambda(\tilde{Q}^*)\|_1 = 1$. In addition, we have

$$\rho(P_{\tilde{Q}^*}) = \sum_{i=1}^M \lambda_i(\tilde{Q}^*) \mathbb{E}[e^{\operatorname{tr}(\tilde{Q}^*\mathrm{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i]. \tag{145}$$

Proof: Since $\psi(\tilde{Q}^*)$ is the right eigenvector associated with the Perron-Frobenius eigenvalue of the matrix $P_{\tilde{Q}^*}$, it holds that

$$\sum_{j=0}^M P(\bar{Q}_j|\bar{Q}_i) e^{\operatorname{tr}(\tilde{Q}^*\bar{Q}_j)} \psi_j(\tilde{Q}^*) = \rho(P_{\tilde{Q}^*}) \psi_i(\tilde{Q}^*) \tag{146}$$

for all $i \in [M]$. From (146) and (143), we can set $\psi(\tilde{Q}^*) = (1, 1, \cdots, 1)^T$ is a right eigenvector associated with the eigenvalue $\rho(P_{\tilde{Q}^*})$ as $\nu \to 0$.

Hence, from Theorem 9, we have

$$Q^* = \lim_{\nu\to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}^*) \mathbb{E}[\boldsymbol{Q}_1 e^{\operatorname{tr}(\tilde{Q}^*\boldsymbol{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i]. \tag{147}$$

Now, since by Theorem 9, it holds that

$$\sum_{j=0}^M \psi_j(\tilde{Q}^*) \lambda_j(\tilde{Q}^*) = 1, \tag{148}$$

so we have

$$\|\lambda(\tilde{Q}^*)\|_1 = 1. \tag{149}$$

Now, since $\lambda(\tilde{Q}^*) := (\lambda_0(\tilde{Q}^*), \lambda_1(\tilde{Q}^*), \cdots, \lambda_M(\tilde{Q}^*))$ is the left (positive) eigenvector associated with $\rho(P_{\tilde{Q}^*})$, it holds that

$$\lambda_j(\tilde{Q}^*) \rho(P_{\tilde{Q}^*}) = \sum_{i=0}^M \lambda_i(\tilde{Q}^*) e^{\operatorname{tr}(\tilde{Q}^*\bar{\mathrm{Q}}_j)} P(\bar{Q}_j|\bar{Q}_i). \tag{150}$$

Then, it follows that

$$\begin{aligned} \rho(P_{\tilde{Q}^*}) &= \sum_{j=0}^M \lambda_j(\tilde{Q}^*) \rho(P_{\tilde{Q}^*}) && (151) \\ &= \sum_{j=0}^M \sum_{i=0}^M \lambda_i(\tilde{Q}^*) e^{\operatorname{tr}(\tilde{Q}^*\bar{\mathrm{Q}}_j)} P(\bar{Q}_j|\bar{Q}_i) && (152) \\ &= \sum_{i=0}^M \lambda_i(\tilde{Q}^*) \sum_{j=1}^M e^{\operatorname{tr}(\tilde{Q}^*\bar{\mathrm{Q}}_j)} P(\bar{Q}_j|\bar{Q}_i) && (153) \\ &= \sum_{i=0}^M \lambda_i(\tilde{Q}^*) \mathbb{E}[e^{\operatorname{tr}(\tilde{Q}^*\mathrm{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i], && (154) \end{aligned}$$

where (151) follows from (149), and (152) follows from (150). ■

Lemma 12. *Under the RS assumption in Definition 10, as $\nu \to 0$, the following hold:*

$$\rho(P_{\tilde{Q}^*}) \to 1 \quad \text{and} \quad \sum_{j=0}^M P(\bar{Q}_j|\bar{Q}_i) e^{\operatorname{tr}(\tilde{Q}^*\bar{Q}_j)} \to 1. \tag{155}$$

Furthermore, it holds that

$$\begin{aligned} &\frac{\partial \log \rho(P_{\tilde{Q}^*})}{\partial \nu}\bigg|_{\nu=0} \\ &= -\frac{\xi}{2}\left(\mathbb{E}[S]\mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left[\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0]\right] + \frac{1}{\eta}\right)\log e + \frac{1}{2}\log\frac{2\pi}{\xi} \\ &\quad + \mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left[\mathbb{E}_S\left\{\int_{\mathbb{R}} p_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta) \right.\right. \\ &\qquad \left.\left.\times \log q_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta) du\right\}\right]. \end{aligned} \tag{156}$$

Proof: By Lemma 5, $\boldsymbol{Q}_0 - \boldsymbol{Q}_1 - \cdots - \boldsymbol{Q}_n$ forms a Markov chain on the state-space $\{\bar{Q}_i\}_{i=0}^M$ defined in Subsection I-D with the transition matrix as (157), where

$$P(\bar{Q}_j|\bar{Q}_i) := \mathbb{P}(\boldsymbol{Q}_1 = \bar{Q}_j | \boldsymbol{Q}_0 = \bar{Q}_i) \tag{158}$$

and $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$ are random (state) matrices at time 0 and 1, respectively.

By [41], we have

$$\begin{aligned} &\min_{i\in[M]} \sum_{j=0}^M P(\boldsymbol{Q}_1 = \bar{Q}_j|\boldsymbol{Q}_0 = \bar{Q}_i) e^{\operatorname{tr}(\tilde{Q}^*\bar{Q}_j)} \le \rho(P_{\tilde{Q}^*}) \\ &\quad \le \max_{i\in[M]} \sum_{j=0}^M P(\boldsymbol{Q}_1 = \bar{Q}_j|\boldsymbol{Q}_0 = \bar{Q}_i) e^{\operatorname{tr}(\tilde{Q}^*\bar{Q}_j)}. \end{aligned} \tag{159}$$

It follows that

$$\begin{aligned} &\min_{i\in[M]} \mathbb{E}\left[e^{\operatorname{tr}(\tilde{Q}^*\boldsymbol{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i\right] \le \rho(P_{\tilde{Q}^*}) \\ &\quad \le \max_{i\in[M]} \mathbb{E}\left[e^{\operatorname{tr}(\tilde{Q}^*\boldsymbol{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i\right]. \end{aligned} \tag{160}$$

First, we show that

$$\lim_{\nu\to 0} \rho(P_{\tilde{Q}^*}) = 1. \tag{161}$$

$$P_{\tilde{Q}} = \begin{bmatrix} P(\bar{Q}_0|\bar{Q}_0)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_0)} & P(\bar{Q}_1|\bar{Q}_0)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_1)} & \cdots & P(\bar{Q}_M|\bar{Q}_0)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_M)} \\ P(\bar{Q}_0|\bar{Q}_1)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_0)} & P(\bar{Q}_1|\bar{Q}_1)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_1)} & \cdots & P(\bar{Q}_M|\bar{Q}_1)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_M)} \\ \vdots & \vdots & \vdots & \vdots \\ P(\bar{Q}_0|\bar{Q}_M)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_0)} & P(\bar{Q}_1|\bar{Q}_M)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_1)} & \cdots & P(\bar{Q}_M|\bar{Q}_M)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_M)} \end{bmatrix}. \tag{157}$$

To show (161), it is enough to show that

$$\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \to 1 \tag{162}$$

for all $i \in [M]$. Indeed, by the definition of $\{\boldsymbol{Q}_n\}_{n=1}^\infty$ in Lemma 5, we have $\boldsymbol{Q}_1 = S_1\boldsymbol{X}_1\boldsymbol{X}_1^T$ where $S_1 \sim P_S$. Hence, we have

$$\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] = \mathbb{E}\left[e^{S_1\boldsymbol{X}_1\tilde{Q}^*\boldsymbol{X}_1^T}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \tag{163}$$

$$= \mathbb{E}\left[e^{S_1\boldsymbol{X}_1\tilde{Q}^*\boldsymbol{X}_1^T}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \tag{164}$$

$$= \mathbb{E}\left[\exp\left(S_1\left[2d\sum_{a=1}^{\nu}X_1^{(0)}X_1^{(a)} + 2f\sum_{1\le a<b\le\nu}X_1^{(a)}X_1^{(b)} + c\big(X_1^{(0)}\big)^2 + g\sum_{a=1}^{\nu}\big(X_1^{(a)}\big)^2\right]\right)\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right], \tag{165}$$

where (165) follows from RS assumption in Definition 10.

Now, the eight parameters (r, m, p, q, f, g) that define Q^* and $\tilde{Q}^*$ are the solution to the joint equations (125) and (126) in Theorem 9. Using (126), it can be shown that [4, Eq. (123)]

$$c = 0, \tag{166}$$

$$d = \frac{1}{2[\sigma^2 + \beta(p-q)]}, \tag{167}$$

$$f = \frac{1 + \beta(r - 2m + q)}{2[\sigma^2 + \beta(p-q)]^2}, \tag{168}$$

$$g = f - d. \tag{169}$$

Now, define $\eta = \frac{2d^2}{f}$ and $\xi = 2d$. In addition, for the simplicity of presentation, let $S := S_1$. Then, by using some algebraic calculation and using the following interesting identity

$$e^{x^2} = \sqrt{\frac{\eta}{2\pi}}\int \exp\left[-\frac{\eta}{2}u^2 + \sqrt{2\eta}xu\right]du, \quad \forall x, \eta, \tag{170}$$

from (165), we have (cf. a similar formula in [4, Eq. (125)]):

$$\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \tag{171}$$

$$= \mathbb{E}\left[\sqrt{\frac{\eta}{2\pi}}\int \exp\left[-\frac{\eta}{2}(u - \sqrt{S}X_1)^2\right]\left[\mathbb{E}_q\left\{\exp\left[-\frac{\xi}{2}u^2 - \frac{\xi}{2}\big(u - \sqrt{S}X\big)^2\right]\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right\}\right]^\nu du\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \tag{172}$$

$$\to \mathbb{E}\left[\sqrt{\frac{\eta}{2\pi}}\int \exp\left[-\frac{\eta}{2}(u - \sqrt{S}X_1)^2\right]du\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \tag{173}$$

$$= \mathbb{E}\big[1\big|\boldsymbol{Q}_0 = \bar{Q}_i\big] \tag{174}$$

$$= 1, \tag{175}$$

where (173) follows from the dominated convergence theorem [43]. Here, as above, we note that the conditional event $\{\boldsymbol{Q}_0 = \bar{Q}_i\}$ only affects the distributions of X and X_1.

Next, we prove that

$$\lim_{\nu\to 0}\frac{\partial\rho(\tilde{Q}^*)}{\partial\nu} = \left(\frac{1}{\log e}\right)\Bigg(-\frac{\xi}{2}\left(\mathbb{E}[S]\mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left[\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0]\right] + \frac{1}{\eta}\right)\log e + \frac{1}{2}\log\frac{2\pi}{\xi} + \mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left[\mathbb{E}_S\left\{\int_{\mathbb{R}} p_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta) \times \log q_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)du\right\}\right]\Bigg). \tag{176}$$

Indeed, at $\nu = 0$, it holds from (172) that

$$\min_{i\in[M]}\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}|\boldsymbol{Q}_0 = \bar{Q}_i\right]\Big|_{\nu=0} = 1 = \max_{i\in[M]}\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}|\boldsymbol{Q}_0 = \bar{Q}_i\right]\Big|_{\nu=0}. \tag{177}$$

Therefore, from (160) and (177), it holds that

$$\rho(P_{\tilde{Q}^*})\big|_{\nu=0} = 1. \tag{178}$$

On the other hand, observe that

$$\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] = \mathbb{E}\left[e^{S\boldsymbol{X}_1\tilde{Q}^*\boldsymbol{X}_1^T}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] \tag{179}$$

$$= \mathbb{E}\left\{\sqrt{\frac{\eta}{2\pi}}\int_{\mathbb{R}}\exp\left[-\frac{\eta}{2}\big(u - \sqrt{S}\mathsf{X}_1\big)^2\right] \times \left[\mathbb{E}_q\left\{\exp\left[-\frac{\xi}{2}u^2 - \frac{\xi}{2}(u - \sqrt{S}\mathsf{X})^2\right]\middle|S\right\}\right]^\nu du\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right\} \tag{180}$$

where (179) follows from (163), (180) follows by using (166)–(169) (see [4, Eq. (125)]). Hence, we have

$$\frac{\partial}{\partial\nu}\mathbb{E}\left[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right] = \left(\frac{1}{\log e}\right)\mathbb{E}\left\{\sqrt{\frac{\eta}{2\pi}}\int_{\mathbb{R}}\exp\left[-\frac{\eta}{2}\big(u - \sqrt{S}\mathsf{X}_1\big)^2\right] \times \left[\mathbb{E}_q\left\{\exp\left[-\frac{\xi}{2}u^2 - \frac{\xi}{2}(u - \sqrt{S}\mathsf{X})^2\right]\middle|S\right\}\right]^\nu \times \log\left(\mathbb{E}_q\left\{\exp\left[-\frac{\xi}{2}u^2 - \frac{\xi}{2}(u - \sqrt{S}\mathsf{X})^2\right]\middle|S\right\}\right)du\middle|\boldsymbol{Q}_0 = \bar{Q}_i\right\}. \tag{181}$$

Hence, given $\mathsf{X}_0 = x_0$ and $S_0 = s_0$, we obtain (187) (see the following page) from (181), which is a constant which does

$$
\begin{aligned}
&\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\mathbb{E}\bigg[e^{\mathrm{tr}(\tilde{Q}^*\boldsymbol{Q}_1)}\bigg|\boldsymbol{Q}_0=\bar{Q}_i\bigg]\\
&=\lim_{\nu\to 0}\bigg(\frac{1}{\log e}\bigg)\mathbb{E}\bigg\{\sqrt{\frac{\eta}{2\pi}}\int_{\mathbb{R}}\exp\bigg[-\frac{\eta}{2}(u-\sqrt{S}\mathsf{X}_1)^2\bigg]\bigg[\mathbb{E}_q\bigg\{\exp\bigg[-\frac{\xi}{2}u^2-\frac{\xi}{2}(u-\sqrt{S}\mathsf{X})^2\bigg]\bigg|S\bigg\}\bigg]^\nu\\
&\qquad\times\log\bigg(\mathbb{E}_q\bigg\{\exp\bigg[-\frac{\xi}{2}u^2-\frac{\xi}{2}(u-\sqrt{S}\mathsf{X})^2\bigg]\bigg|S\bigg\}\bigg)du\bigg|\boldsymbol{Q}_0=\bar{Q}_i\bigg\} &&(182)\\
&=\bigg(\frac{1}{\log e}\bigg)\mathbb{E}\bigg\{\sqrt{\frac{\eta}{2\pi}}\int_{\mathbb{R}}\exp\bigg[-\frac{\eta}{2}(u-\sqrt{S}\mathsf{X}_1)^2\bigg]\log\bigg(\mathbb{E}_q\bigg\{\exp\bigg[-\frac{\xi}{2}u^2-\frac{\xi}{2}(u-\sqrt{S}\mathsf{X})^2\bigg]\bigg|S\bigg\}\bigg)du\bigg|\mathsf{X}_0=x_0\bigg\} &&(183)\\
&=\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\mathbb{E}\bigg\{\sqrt{\frac{\eta}{2\pi}}\int_{\mathbb{R}}\exp\bigg[-\frac{\eta}{2}(u-\sqrt{S}\mathsf{X}_1)^2\bigg]u^2du\bigg|\mathsf{X}_0=x_0\bigg\}\\
&\qquad+\mathbb{E}\bigg\{\int_{\mathbb{R}}p_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u|\mathsf{X}_0,\mathsf{X}_1,S;\eta)\log\bigg(\mathbb{E}_q\bigg\{\sqrt{\frac{2\pi}{\xi}}q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u|\mathsf{X}_0,\mathsf{X}_1,S;\eta)\bigg|S\bigg\}\bigg)du\bigg|\mathsf{X}_0=x_0\bigg\}\bigg) &&(184)\\
&=\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\bigg(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0=x_0]+\frac{1}{\eta}\bigg)\log e+\frac{1}{2}\log\frac{2\pi}{\xi}\\
&\qquad+\mathbb{E}_S\bigg\{\int_{\mathbb{R}}\mathbb{E}_{\pi(x_0,\cdot)}\bigg[p_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u|x_0,\mathsf{X}_1,S;\eta)\bigg|S\bigg]\log\bigg(\mathbb{E}_{\tilde{\pi}(x_0,\cdot)}\bigg[q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u|x_0,\mathsf{X}_1,S;\eta)\bigg|S\bigg]\bigg)du\bigg|\mathsf{X}_0=x_0\bigg\}\bigg) &&(185)\\
&=\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\bigg(\mathbb{E}[S\mathsf{X}_1^2|\mathsf{X}_0=x_0]+\frac{1}{\eta}\bigg)\log e+\frac{1}{2}\log\frac{2\pi}{\xi}\\
&\qquad+\mathbb{E}_S\bigg\{\int_{\mathbb{R}}\mathbb{E}_{\pi(x_0,\cdot)}\big[p_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u|\mathsf{X}_0,\mathsf{X}_1,S;\eta)\big]\log\bigg(\mathbb{E}_{\tilde{\pi}(x_0,\cdot)}\bigg\{q_{U|\mathsf{X}_0,\mathsf{X}_1,S;\eta}(u|\mathsf{X}_0,\mathsf{X}_1,S;\eta)\bigg|S\bigg\}\bigg)du\bigg|\mathsf{X}_0=x_0\bigg\}\bigg) &&(186)\\
&=\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\bigg(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0=x_0]+\frac{1}{\eta}\bigg)\log e+\frac{1}{2}\log\frac{2\pi}{\xi}\\
&\qquad+\mathbb{E}_S\bigg\{\int_{\mathbb{R}}p_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)\log q_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)du\bigg\}\bigg). &&(187)
\end{aligned}
$$

not depend on $\bar{Q}_i$, where (183) follows from the dominated convergence theorem [43]. Here, we note that the conditional event $\{\boldsymbol{Q}_0=\bar{Q}_i\}$ only affects the distribution of X and X_1.

Now, from Lemma 11, it holds that

$$
\begin{aligned}
\frac{\partial\rho(P_{\tilde{Q}^*})}{\partial\nu}\bigg|_{\nu=0}&=\lim_{\nu\to 0}\frac{\rho(P_{\tilde{Q}^*})\big|_\nu-1}{\nu} &&(188)\\
&=\lim_{\nu\to 0}\sum_{i=1}^M\lambda_i(\tilde{Q}^*)\bigg(\frac{\mathbb{E}[e^{\mathrm{tr}(\tilde{Q}^*\mathrm{Q}_1)}|\boldsymbol{Q}_0=\bar{Q}_i]-1}{\nu}\bigg) &&(189)
\end{aligned}
$$

where (188) follows from (178).

Finally, as $\nu\to 0$, it holds that $\tilde{Q}^*\to c=0$ by (166) and (142) of Definition 10. Therefore, we have $P_{\tilde{Q}^*}\to P_S\otimes P_\pi$ and $M\to\big|\{sx^2:(x,s)\in\mathcal{X}\times\mathcal{S}\}\big|:=M_0$, where $\otimes$ is denoted as the Kronecker product. It follows that for each fixed $S=s$, $\lambda(\tilde{Q})\to\tilde{\lambda}^{(\pi)}$ where $\tilde{\lambda}^{(\pi)}$ is the left Perron-Frobenius eigenvector of the stochastic matrix $P_S\otimes P_\pi$ such that $\|\lambda^{(\pi)}\|_1=1$. By [50, p. 7], the left Perron-Frobenius eigenvector exists, and it is unique up to a positive scaling factor, so $\tilde{\lambda}^{(\pi)}$ exists uniquely.

Let $\lambda^{(\pi)}$ be the marginal distribution of $\tilde{\lambda}^{(\pi)}$. Then, from (187) and (189), we obtain

$$
\begin{aligned}
&\frac{\partial\rho(P_{\tilde{Q}^*})}{\partial\nu}\bigg|_{\nu=0}\\
&=\sum_{s\in\mathcal{S}}\sum_{x_0\in\mathcal{X}_0}\tilde{\lambda}_{s,x_0}^{(\pi)}\\
&\quad\times\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\bigg(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0=x_0]+\frac{1}{\eta}\bigg)\log e\\
&\qquad+\frac{1}{2}\log\frac{2\pi}{\xi}+\mathbb{E}_S\bigg\{\int_{\mathbb{R}}p_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)\\
&\qquad\times\log q_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)du\bigg\}\bigg) &&(190)\\
&=\sum_{x_0\in\mathcal{X}_0}\lambda_{x_0}^{(\pi)}\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\bigg(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0=x_0]+\frac{1}{\eta}\bigg)\log e\\
&\qquad+\frac{1}{2}\log\frac{2\pi}{\xi}+\mathbb{E}_S\bigg\{\int_{\mathbb{R}}p_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)\\
&\qquad\times\log q_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)du\bigg\}\bigg) &&(191)\\
&=\bigg(\frac{1}{\log e}\bigg)\bigg(-\frac{\xi}{2}\bigg(\mathbb{E}[S]\mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\Big[\mathbb{E}[\mathsf{X}_1^2|\mathsf{X}_0]\Big]+\frac{1}{\eta}\bigg)\log e\\
&\qquad+\frac{1}{2}\log\frac{2\pi}{\xi}+\mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\bigg[\mathbb{E}_S\bigg\{\int_{\mathbb{R}}p_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)\\
&\qquad\times\log q_{U|\mathsf{X}_0,S;\eta}(u|x_0,S;\eta)du\bigg\}\bigg]\bigg). &&(192)
\end{aligned}
$$

This concludes our proof of Lemma 12. ■

B. Proofs of Claims

1) Proof of Claim 1: Recall the definitions of $\{\bar{Q}_i\}_{i=1}^M$ in Subsection I-D. From Lemma 11, it holds that

$$Q^* = \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}^*) \mathbb{E}[\boldsymbol{Q}_1 e^{\mathrm{tr}(\tilde{Q}^* \boldsymbol{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i], \tag{193}$$

where $\|\lambda(\tilde{Q}^*)\|_1 = 1$ and all its components are positive.

By Lemma 4, we have $\boldsymbol{Q}_1 = S_1 \boldsymbol{X}_1 \boldsymbol{X}_1^T$ and $\boldsymbol{Q}_0 = S_0 \boldsymbol{X}_0 \boldsymbol{X}_0^T$ where $\boldsymbol{X}_1 := (X_1^{(0)}, X_1^{(1)}, \cdots, X_1^{(\nu)})^T$ and $\boldsymbol{X}_0 := (X_0^{(0)}, X_0^{(1)}, \cdots, X_0^{(\nu)})^T$. It follows that for any $\tilde{Q} \in \mathcal{Q}$ and $\bar{Q}_i \in \mathcal{Q}$, we have

$$\hat{Q}_i(\tilde{Q}) := \mathbb{E}[\boldsymbol{Q}_1 e^{\mathrm{tr}(\tilde{Q}\boldsymbol{Q}_1)} | \boldsymbol{Q}_0 = \bar{Q}_i] \tag{194}$$

$$= \mathbb{E}\left[S_1 \boldsymbol{X}_1 \boldsymbol{X}_1^T \exp\left[\boldsymbol{X}_1^T \tilde{Q} \boldsymbol{X}_1\right] \middle| S_0 = s_i, \boldsymbol{X}_0 = x_i \right] \tag{195}$$

$$= \mathbb{E}\left[S_1 \boldsymbol{X}_1 \boldsymbol{X}_1^T \exp\left[\boldsymbol{X}_1^T \tilde{Q} \boldsymbol{X}_1\right] \middle| \boldsymbol{X}_0 = x_i \right] \tag{196}$$

for some $s_i \in \mathcal{S}$ and $x_i \in \mathcal{X}^{\nu+1}$ such that $s_i x_i x_i^T = \bar{Q}_i$, where (195) follows from the uniqueness of the x_i and s_i by the definition of $\mathcal{Q}$ in (3), and (196) follows from the fact that S_0 is independent of $\boldsymbol{X}_1, \boldsymbol{X}_0$.

This means that for each fixed $i \in [M]$, $\hat{Q}_i^{(a,b)}(\tilde{Q})$ is in the same form as [4, Eq. (127)] for each $(a,b) \in [\nu+1] \times [\nu+1]$. Hence, by setting $S := S_1 \sim P_S$ as above, we have

$$\hat{Q}_i^{(0,1)}(\tilde{Q})$$

$$= \mathbb{E}\left[S X_1^{(0)} X_1^{(1)} \exp\left[\boldsymbol{X}_1^T \tilde{Q} \boldsymbol{X}_1\right] \middle| \boldsymbol{X}_0 = x_i \right] \tag{197}$$

$$= \mathbb{E}\left[S \mathsf{X}_1 \langle \mathsf{X} | \boldsymbol{X}_0 = x_i \rangle_q \middle| \boldsymbol{X}_0 = x_i \right], \tag{198}$$

where (198) follows from [4, Eq. (131)].

Similarly, we also have

$$r_i := \hat{Q}_i^{(0,0)} = \mathbb{E}\big[S \big| \boldsymbol{X}_0 = x_i\big], \tag{199}$$

$$m_i := \hat{Q}_i^{(0,1)} = \mathbb{E}\big[S \mathsf{X}_1 \langle \mathsf{X} | \boldsymbol{X}_0 = x_i \rangle_q \big| \boldsymbol{X}_0 = x_i\big], \tag{200}$$

$$p_i := \hat{Q}_i^{(1,1)} = \mathbb{E}\big[S \mathsf{X}^2 \big| \boldsymbol{X}_0 = x_i\big], \tag{201}$$

$$q_i := \hat{Q}_i^{(1,2)} = \mathbb{E}\big[S \langle \mathsf{X} | \boldsymbol{X}_0 = x_i \rangle_q^2 \big| \boldsymbol{X}_0 = x_i\big], \tag{202}$$

for all $i \in [M]$. Since $\boldsymbol{X}^{(a)} \sim q_{\boldsymbol{X}}$ for all $a = 1, 2, \cdots, \nu$ and mutually independent to each other, it follows from (199)–(202) that $\hat{Q}_i(\tilde{Q})$ has the RS form as defined in Lemma 10, i.e.,

$$\hat{Q}_i(\tilde{Q}) = \begin{bmatrix} r_i & m_i & m_i & \cdots & m_i \\ m_i & p_i & q_i & \cdots & q_i \\ m_i & q_i & p_i & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & q_i \\ m_i & q_i & \cdots & q_i & p_i \end{bmatrix}. \tag{203}$$

for all $i \in [M]$.

It follows from (193) and (194) that

$$Q^*(\tilde{Q}) = \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \hat{Q}_i(\tilde{Q}) \tag{204}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \begin{bmatrix} r_i & m_i & m_i & \cdots & m_i \\ m_i & p_i & q_i & \cdots & q_i \\ m_i & q_i & p_i & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & q_i \\ m_i & q_i & \cdots & q_i & p_i \end{bmatrix}. \tag{205}$$

Hence, from the RS assumption in Definition 10 and (205), we obtain

$$r = \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) r_i \tag{206}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \big| \boldsymbol{X}_0 = x_i\big] \tag{207}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \big| \mathsf{X}_0 = x_i^{(0)}\big], \tag{208}$$

where $x_i^{(0)}$ is the first element of the vector x_i. In addition, we also have

$$m = \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) m_i \tag{209}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \mathsf{X}_1 \langle \mathsf{X} | \boldsymbol{X}_0 = x_i \rangle_q \big| \boldsymbol{X}_0 = x_i\big] \tag{210}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \mathsf{X}_1 \langle \mathsf{X} | \mathsf{X}_0 = x_i^{(0)} \rangle_q \big| \mathsf{X}_0 = x_i^{(0)}\big], \tag{211}$$

$$p = \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) p_i \tag{212}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \mathsf{X}^2 \big| \boldsymbol{X}_0 = x_i\big] \tag{213}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \mathsf{X}^2 \big| \mathsf{X}_0 = x_i^{(0)}\big], \tag{214}$$

$$q = \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) q_i \tag{215}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \langle \mathsf{X} | \boldsymbol{X}_0 = x_i \rangle_q^2 \big| \boldsymbol{X}_0 = x_i\big] \tag{216}$$

$$= \lim_{\nu \to 0} \sum_{i=0}^M \lambda_i(\tilde{Q}) \mathbb{E}\big[S \langle \mathsf{X} | \mathsf{X}_0 = x_i^{(0)} \rangle_q^2 \big| \mathsf{X}_0 = x_i^{(0)}\big]. \tag{217}$$

From these facts, we obtain

$$r-2m+q=\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\bigg[S\bigg(\mathsf{X}_1^2-2\mathsf{X}_1\langle\mathsf{X}|\mathsf{X}_0=x_i^{(0)}+\langle\mathsf{X}|\mathsf{X}_0=x_i^{(0)}\rangle_q^2\bigg)\bigg|\mathsf{X}_0=x_i^{(0)}\bigg] \tag{218}$$

$$=\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\bigg[S\bigg(\mathsf{X}_1-\langle\mathsf{X}|\mathsf{X}_0=x_i^{(0)}\rangle_q\bigg)^2\bigg|\mathsf{X}_0=x_i^{(0)}\bigg], \tag{219}$$

and similarly,

$$p-q=\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\bigg[S\bigg(\mathsf{X}-\langle\mathsf{X}|\mathsf{X}_0=x_i^{(0)}\rangle_q\bigg)^2\bigg|\mathsf{X}_0=x_i^{(0)}\bigg]. \tag{220}$$

On the other hand, from (166)–(169), we also have

$$r-2m+q=\frac{1}{\beta}\bigg(\frac{1}{\eta}-1\bigg), \tag{221}$$

$$p-q=\frac{1}{\beta}\bigg(\frac{1}{\xi}-\sigma^2\bigg). \tag{222}$$

From (219)–(222), (η,ξ) is a solution of the following equation system:

$$\eta^{-1}=1+\beta\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\times\mathbb{E}\bigg[S\bigg(\mathsf{X}_1-\langle\mathsf{X}|\mathsf{X}_0=x_i^{(0)}\rangle_q\bigg)^2\bigg|\mathsf{X}_0=x_i^{(0)}\bigg], \tag{223}$$

$$=1+\beta\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\bigg[S\mathcal{E}(S;\eta,\xi|\mathsf{X}_0=x_i^{(0)})\bigg], \tag{224}$$

$$\xi^{-1}=\sigma^2+\beta\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\times\mathbb{E}\bigg[S\bigg(\mathsf{X}-\langle\mathsf{X}|\mathsf{X}_0=x_i^{(0)}\rangle_q\bigg)^2\bigg|\mathsf{X}_0=x_i^{(0)}\bigg] \tag{225}$$

$$=\sigma^2+\beta\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\bigg[S\mathcal{V}(S;\eta,\xi|\mathsf{X}_0=x_i^{(0)})\bigg]. \tag{226}$$

Now, from (107) in Lemma 4 and RS assumption on Definition 10, we obtain

$$G^{(\nu)}(Q^*)=-\frac{\nu}{2}\log(2\pi\sigma^2)-\frac{\nu-1}{2}\log\bigg[1+\frac{\beta}{\sigma^2}(p-q)\bigg]-\frac{1}{2}\log\bigg[1+\frac{\beta}{\sigma^2}(p-q)+\frac{\nu}{\sigma^2}\bigg(1+\beta(r-2m+q)\bigg)\bigg]. \tag{227}$$

In addition, we also have

$$I^{(\nu)}(Q^*)$$
$$=\mathrm{tr}(\tilde{Q}^*Q^*)-\log\rho(P_{\tilde{Q}^*}) \tag{228}$$
$$=\mathrm{tr}(\tilde{Q}^*Q^*)-\log\rho(P_{\tilde{Q}^*}), \tag{229}$$
$$=rc+\nu pg+2\nu md+\nu(\nu-1)qf-\log\rho(P_{\tilde{Q}^*}), \tag{230}$$

where (229) follows from Lemma 12, and (230) follows from assumptions Q^* and $\tilde{Q}^*$ in Definition 10.

Now, by the RS assumption, the eight parameters (r,m,p,q,c,d,f,g) have zero derivatives with respect to ν as $\nu\to 0$ [4, p.1999]. Let $\lambda^{(\pi)}$ is the left Perron-Frobenius eigenvector of the stochastic matrix P_π such that $\|\lambda^{(\pi)}\|_1=1$, which is the stationary distribution of the stochastic matrix. By choosing the initial state at the state that the limit distribution of the Markov process $\{X_n\}_{n=1}^\infty$ converges to the stationary distribution. Then, from Theorem 8, we have

$$\mathcal{F}_q$$
$$=-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}(\beta^{-1}G^{(\nu)}(Q^*)-I^{(\nu)}(Q^*)) \tag{231}$$

$$=\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\bigg(rc+\nu pg+2\nu md+\nu(\nu-1)qf-\beta^{-1}\bigg(-\frac{\nu}{2}\log(2\pi\sigma^2)-\frac{\nu-1}{2}\log\bigg[1+\frac{\beta}{\sigma^2}(p-q)\bigg]-\frac{1}{2}\log\bigg[1+\frac{\beta}{\sigma^2}(p-q)+\frac{\nu}{\sigma^2}\bigg(1+\beta(r-2m+q)\bigg)\bigg]\bigg)\bigg)-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\rho(P_{\tilde{Q}^*}) \tag{232}$$

$$=pg+2md-qf+\beta^{-1}\bigg[\frac{1}{2}\log(2\pi\sigma^2)+\frac{1}{2}\log\bigg(1+\frac{\beta}{\sigma^2}(p-q)\bigg)+\frac{1+\beta(r-2m+q)}{2\sigma^2(1+\frac{\beta}{\sigma^2}(p-q))}\log e\bigg]-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\rho(P_{\tilde{Q}^*}) \tag{233}$$

$$=p(f-d)+2md-qf+\beta^{-1}\bigg[\frac{1}{2}\log(2\pi\sigma^2)+\frac{1}{2}\log\bigg(1+\frac{\beta}{\sigma^2}(p-q)\bigg)+\frac{1+\beta(r-2m+q)}{2\sigma^2(1+\frac{\beta}{\sigma^2}(p-q))}\log e\bigg]-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\rho(P_{\tilde{Q}^*}) \tag{234}$$

$$=(p-q)f-pd+2md+\beta^{-1}\bigg[\frac{1}{2}\log(2\pi\sigma^2)+\frac{1}{2}\log\bigg(1+\frac{\beta}{\sigma^2}(p-q)\bigg)+\frac{1+\beta(r-2m+q)}{2\sigma^2(1+\frac{\beta}{\sigma^2}(p-q))}\log e\bigg]-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\rho(P_{\tilde{Q}^*}) \tag{235}$$

$$=\frac{1}{\beta}\bigg(\frac{1}{\xi}-\sigma^2\bigg)\frac{\xi^2}{2\eta}-\frac{p\xi}{2}+\xi m+\frac{1}{\beta}\bigg[\frac{1}{2}\log(2\pi\sigma^2)-\frac{1}{2}\log\big(\xi\sigma^2\big)+\frac{\xi^2}{2\eta}\log e\bigg]-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\rho(P_{\tilde{Q}^*}) \tag{236}$$

$$=\xi m-\frac{p\xi}{2}+\frac{1}{\beta}\bigg(\frac{1}{\xi}-\sigma^2\bigg)\frac{\xi^2}{2\eta}-\frac{1}{2\beta}\log\xi+\frac{1}{2\beta}\log(2\pi)+\frac{\xi^2}{2\beta\eta}\log e-\lim_{\nu\to 0}\frac{\partial}{\partial\nu}\log\rho(P_{\tilde{Q}^*}) \tag{237}$$

$$=\xi m-\frac{p\xi}{2}+\frac{1}{\beta}\bigg(\frac{1}{\xi}-\sigma^2\bigg)\frac{\xi^2}{2\eta}-\frac{1}{2\beta}\log\xi+\frac{1}{2\beta}\log(2\pi)$$

$$+ \frac{\xi^2}{2\beta\eta}\log e + \frac{\xi}{2}\left(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2] + \frac{1}{\eta}\right)\log e - \frac{1}{2}\log\frac{2\pi}{\xi}$$
$$- \mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left\{\mathbb{E}_S\left\{\int_{\mathbb{R}} p_{U|\mathsf{X}_0,S;\eta}(u|X_0,S;\eta)\right\}\right.$$
$$\left.\times \log\left(q_{U|\mathsf{X}_0,S;\eta}(u|X_0,S;\eta)\right)du\right\}\Bigg\}\Bigg) \tag{238}$$
$$= \xi \lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\big[S\mathsf{X}_1\langle \mathsf{X}|\mathsf{X}_0 = x_i^{(0)}\rangle_q\big|\mathsf{X}_0 = x_i^{(0)}\big]$$
$$- \frac{\xi}{2}\lim_{\nu\to 0}\sum_{i=0}^{M}\lambda_i(\tilde{Q})\mathbb{E}\big[S\mathsf{X}^2\big|\mathsf{X}_0 = x_i^{(0)}\big]$$
$$+ \frac{1}{\beta}\left(\frac{1}{\xi} - \sigma^2\right)\frac{\xi^2}{2\eta} - \frac{1}{2\beta}\log\xi + \frac{1}{2\beta}\log(2\pi)$$
$$+ \frac{\xi^2}{2\beta\eta}\log e + \frac{\xi}{2}\left(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2] + \frac{1}{\eta}\right)\log e - \frac{1}{2}\log\frac{2\pi}{\xi}$$
$$- \mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left\{\mathbb{E}_S\left\{\int_{\mathbb{R}} p_{U|\mathsf{X}_0,S;\eta}(u|X_0,S;\eta)\right.\right.$$
$$\left.\left.\times \log\left(q_{U|\mathsf{X}_0,S;\eta}(u|X_0,S;\eta)\right)du\right\}\right\} \tag{239}$$
$$= \xi\mathbb{E}_{X_0\sim\lambda^{(\pi)}}\Big[\mathbb{E}\big[S\mathsf{X}_1\langle \mathsf{X}|\mathsf{X}_0\rangle_q\big|\mathsf{X}_0\big]\Big]$$
$$- \frac{\xi}{2}\mathbb{E}_{X_0\sim\lambda^{(\pi)}}\Big[\mathbb{E}\big[S\mathsf{X}^2\big|\mathsf{X}_0\big]\Big] + \frac{1}{\beta}\left(\frac{1}{\xi} - \sigma^2\right)\frac{\xi^2}{2\eta}$$
$$- \frac{1}{2\beta}\log\xi + \frac{1}{2\beta}\log(2\pi) + \frac{\xi^2}{2\beta\eta}\log e$$
$$+ \frac{\xi}{2}\left(\mathbb{E}[S]\mathbb{E}[\mathsf{X}_1^2] + \frac{1}{\eta}\right)\log e - \frac{1}{2}\log\frac{2\pi}{\xi}$$
$$- \mathbb{E}_{\mathsf{X}_0\sim\lambda^{(\pi)}}\left\{\mathbb{E}_S\left\{\int_{\mathbb{R}} p_{U|\mathsf{X}_0,S;\eta}(u|X_0,S;\eta)\right.\right.$$
$$\left.\left.\times \log\left(q_{U|\mathsf{X}_0,S;\eta}(u|X_0,S;\eta)\right)du\right\}\right\} \tag{240}$$
$$= \sum_{x_0}\lambda_{x_0}^{(\pi)}\mathcal{G}(x_0), \tag{241}$$

where (233) follows from Lemma 12, (234) follows from (166)–(169), (236) follows from (221) and (222), (238) follows from Lemma 12, (239) follows from (211) and (214), (240) follows from $\lambda(\tilde{Q}) \to \lambda^{(\pi)}$ since $P_{\tilde{Q}} \to P_\pi$ as $\nu \to 0$ where $\lambda^{(\pi)}$ is the left Perron-Frobenius eigenvector of the stochastic matrix P_π such that $\|\lambda^{(\pi)}\|_1 = 1$, which is the stationary distribution of the Markov chain $\{X_n\}_{n=1}^\infty$[8], and (241) follows from [4, Sect. IV].

Hence, we obtain (38) from (224), (226), and (241).

Finally, (39) is an direct application of [3, Prop. 5].

2) Proof of Claim 2: The result in (40) can be obtained by using the same ideas as in the proof of Theorem 8, [4, Sec. IV-B], and the facts in (193), (196), and (241). The detailed proof can be found in [47, Appendix C].

Now, observe that by using the MMSE decoder defined in Section II-A, we have

$$\mathbb{E}\big[\|\boldsymbol{X} - [\boldsymbol{X}]\|_2^2\big] = \sum_{k=1}^{n}\mathbb{E}\big[\mathsf{X}_k^2\big] - \sum_{k=1}^{n}\mathbb{E}\big[[\mathsf{X}_k]^2\big], \tag{242}$$

where (242) follows from

$$[\mathsf{X}_k] = \mathbb{E}_p\big[\mathsf{X}_k\big|\boldsymbol{Y},\boldsymbol{\Phi}\big], \tag{243}$$

which is drawn from (10).

Now, by (40), we have as $n \to \infty$,

$$\mathbb{E}\big[[\mathsf{X}_k]^2\big] = \sum_{x_0\in\mathcal{X}}\lambda_{x_0}^{(\pi)}\mathbb{E}\big[\langle \mathsf{X}_1|\mathsf{X}_0 = x_0\rangle^2\big], \quad \forall k \in [n]. \tag{244}$$

In addition, for all $k \in [n]$, we also have

$$\mathbb{E}\big[\mathsf{X}_k^2\big] = \mathbb{E}\big[\mathbb{E}\big[\mathsf{X}_k^2|\mathsf{X}_{k-1}\big]\big] \tag{245}$$
$$= \mathbb{E}\big[\mathbb{E}\big[\mathsf{X}_1^2|\mathsf{X}_0\big]\big] \tag{246}$$
$$= \mathbb{E}[\mathsf{X}_1^2], \tag{247}$$

where (245) follows from the tower property [48], and (246) follows from the time-homogeneity of Markov process $\{X_n\}_{n=1}^\infty$.

From (242), (244), and (247), as $n \to \infty$, we have

$$\mathbb{E}\big[\|\boldsymbol{X} - [\boldsymbol{X}]\|_2^2\big] = n\bigg(\mathbb{E}[\mathsf{X}_1^2] - \sum_{x_0\in\mathcal{X}}\lambda_{x_0}^{(\pi)}\mathbb{E}\big[\langle \mathsf{X}_1|\mathsf{X}_0 = x_0\rangle^2\big]\bigg), \tag{248}$$

which leads to (41).

3) Proof of Claim 3: First, for any $n \geq 2$, by using Markov chains such as $\Upsilon_n - \Upsilon_{n-1} - (X_{n-1}, \{X_k, \Upsilon_k\}_{k=1}^{n-2})$ and $X_n - \Upsilon_n - (\{X_k, \Upsilon_k\}_{k=1}^{n-1})$, we can easily show that $\{(X_n, \Upsilon_n)\}_{n=1}^\infty$ forms a Markov chain with states on $\mathcal{X}\times\mathcal{S}_\Upsilon$. Hence, (58), (59), and (60) are direct results of Claim 1 and Claim 2.

Appendix A
Proof of Lemma 7

Consider the homogeneous Markov chain with states in the set $\mathcal{Q}$ as mentioned in Lemma 5. This Markov chains have M states $\bar{Q}_0, \bar{Q}_1, \cdots, \bar{Q}_M$ where $M = |\mathcal{Q}| - 1$. Recall the definition of $P_{\tilde{Q}}$ in (157).

Then, $P_{\tilde{Q}}$ is an irreducible non-negative matrix [41]. Let $\rho(P_{\tilde{Q}})$ denote the Perron-Frobenious eigenvalue of the non-negative irreducible matrix $P_{\tilde{Q}}$. It follows from [50, p. 7] and [50, Lemma 3.1] that

$$\frac{\partial\rho(P_{\tilde{Q}})}{\partial P_{\tilde{Q}}} = \lambda(\tilde{Q})\psi(\tilde{Q})^T \tag{249}$$

[8] By [50, p. 7], the left Perron-Frobenius eigenvector exists, and it is unique up to a positive scaling factor, so $\lambda^{(\pi)}$ exists uniquely.

where $\lambda(\tilde{Q})$ and $\psi(\tilde{Q})$ are left and right eigenvectors associated with the eigenvalue $\rho(P_{\tilde{Q}})$ which are normalized such that $\lambda(\tilde{Q})^T\psi(\tilde{Q})=1$. Hence, we have

$$\frac{\partial\rho(P_{\tilde{Q}})}{\partial\tilde{Q}}(\tilde{Q})$$

$$=\frac{\partial\rho(P_{\tilde{Q}})}{\partial P_{\tilde{Q}}}o_{\rm tr}\frac{\partial P_{\tilde{Q}}}{\partial\tilde{Q}} \tag{250}$$

$$=\lambda(\tilde{Q})\psi(\tilde{Q})^T o_{\rm tr} \begin{bmatrix} \bar{Q}_0 P(\bar{Q}_0|\bar{Q}_0)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_0)} & \cdots & \bar{Q}_M P(\bar{Q}_M|\bar{Q}_0)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_M)} \\ \bar{Q}_0 P(\bar{Q}_0|\bar{Q}_1)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_0)} & \cdots & \bar{Q}_M P(\bar{Q}_M|\bar{Q}_1)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_M)} \\ \vdots & \vdots & \vdots \\ \bar{Q}_0 P(\bar{Q}_0|\bar{Q}_M)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_0)} & \cdots & \bar{Q}_M P(\bar{Q}_M|\bar{Q}_M)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_M)} \end{bmatrix} \tag{251}$$

$$=\sum_{i=0}^M\sum_{j=0}^M \lambda_i(\tilde{Q})\psi_j(\tilde{Q})\bar{Q}_j P(\bar{Q}_j|\bar{Q}_i)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_j)} \tag{252}$$

$$=\sum_{i=0}^M \lambda_i(\tilde{Q})\sum_{j=0}^M \psi_j(\tilde{Q})\bar{Q}_j P(\bar{Q}_j|\bar{Q}_i)e^{\mathrm{tr}(\tilde{Q}\bar{Q}_j)}. \tag{253}$$

Now, by the chain rule, we also have

$$\frac{\partial\log\rho(P_{\tilde{Q}})}{\partial\tilde{Q}}(\tilde{Q})=\frac{1}{\rho(P_{\tilde{Q}})}\frac{\partial P_{\tilde{Q}}}{\partial\tilde{Q}}(\tilde{Q}). \tag{254}$$

Hence, we obtain (117) from (253) and (254).

Appendix B
Extensions to Markov chains on a general Polish space in $\mathbb{R}$

In this section, we sketch what we should change in our analysis when working with a Markov chains on a general Polish space in $\mathbb{R}$.

- As the spectral method (Paulin), we define an associated linear operator $\boldsymbol{\pi}$ on L_2 to a the Markov kernel $\pi(x,y)$ such that

$$\boldsymbol{\pi}(f)(x):=\int_{\mathcal{S}}\pi(x,y)f(y)dy. \tag{255}$$

  We call $f(\cdot)$ is an eigenvector of π associated with an eigenvalue λ if and only if $\boldsymbol{\pi}(f)(x)=\lambda f(x)$ for all $x\in\mathcal{S}$. The existence of such λ and f is guaranteed (for example, let $f(y)=1/\|f\|,\forall y\in\mathcal{S}$ and $\lambda=1$). Define S_2 be the set of eigenvalues of π. The Perron-Frobenius eigenvalue is defined as the supremum of all elements in this set[9].
- Then, we show that for every positive function $h:\mathbb{S}\to\mathbb{R}_+$ and Markov chain $\{Z_n\}_{n=1}^\infty$ on an arbitrary space $\mathcal{V}$ with stochastic kernel $Q(x,y)$, the following holds:

$$\lim_{n\to\infty}\frac{1}{n}\log\left[\int_{\mathcal{V}}Q^n(x,y)h(y)dy\right]=\log\rho(Q),\quad\forall x\in\mathcal{V}, \tag{256}$$

  where $\rho(Q)$ is the Perron-Frobenius eigenvalue of Q.

[9]Since the linear operator is continuous (bounded), the set of eigenvalues is bounded.

- Show that $\boldsymbol{T}_n=\frac{1}{n}\sum_{k=1}^n\boldsymbol{Q}_k$ satisfies the large deviation bounds with rate $I(Q)=\sup_{\tilde{Q}}(\mathrm{tr}(\tilde{Q}Q)-\log\rho(P_{\tilde{Q}})$, where $\rho(P_{\tilde{Q}})$ is the Perron-Frobenius eigenvalue of the Markov chain $\boldsymbol{Q}_0-\boldsymbol{Q}_1\cdots-\boldsymbol{Q}_n$.
- By Varadhan theorem on Polish space [49], we can show that Lemma 6 still holds, i.e.,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\big[e^{nF(\boldsymbol{T}_n)}\big]$$

$$=\lim_{n\to\infty}\frac{1}{n}\log\int e^{nF(Q)}dP_n(Q) \tag{257}$$

$$=\sup_Q\Big[F(Q)-I(Q)\Big] \tag{258}$$

  for any bounded continuous function $F:\mathcal{Q}\to\mathbb{R}$. The main difference is now $|\mathcal{Q}|$ is unbounded or $M\to\infty$.
- From (258), by applying for a specific function F, we obtain Theorem 8.
- The rest is an optimization problem and the same arguments as previous section still work.

Acknowledgements

The author is extremely grateful to Prof. Ramji Venkataramanan, the University of Cambridge, for many suggestions to improve the paper. The author also would like to thank the editor and reviewers for their suggestions to improve the paper.

References

[1] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Phys. F: Metal Physics*, vol. 5, pp. 965–974, 1975.
[2] A. Bereyhi, R. Müller, and H. S. Baldes, "Statistical mechanics of MAP estimation: General replica ansatz," *IEEE Trans. on Inform. Th.*, vol. 65, no. 12, pp. 7896–7934, Dec. 2019.
[3] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. on Inform. Th.*, vol. 48, no. 11, pp. 2888–2909, 2002.
[4] Dongning Guo and S. Verdu, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.
[5] R. R. Müller, "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Trans. Signal Process.*, vol. 51, pp. 2821–2828, 2003.
[6] Y. Kabashima, T. Wadayama, and T. Tanaka, "Typical reconstruction limit of compressed sensing based on l_p-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, 2009.
[7] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 52–59, 2009.
[8] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Trans. on Inform. Th.*, vol. 58, no. 3, pp. 1902–1923, March 2012.
[9] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, pp. 3065–3092, 2012.
[10] ——, "Compressed sensing phase transitions: Rigorous bounds versus replica predictions," *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2012.
[11] A. M. Tulino, G. Caire, S. Verdú, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4243–4271, 2013.
[12] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," *2006 IEEE Information Theory Workshop - ITW '06 Chengdu*, pp. 194–198, 2006.

[13] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," *2006 IEEE Information Theory Workshop - ITW '06 Punta del Este*, pp. 160–164, 2006.

[14] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 269–280, 2010.

[15] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random CDMA systems," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5590–5613, 2010.

[16] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual information and optimality of approximate message-passing in random linear estimation," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4270–4303, 2020.

[17] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *The 48th Annual Allerton Conference on Communications and Computation*, 2010, pp. 347–353.

[18] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012.

[19] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Transactions on Information Theory*, vol. 59, pp. 7434–7464, 2011.

[20] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for random linear estimation with gaussian matrices is exact," *IEEE Trans. on Inform. Th.*, vol. 65, no. 4, pp. 2252–2283, April 2019.

[21] Dongning Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[22] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5451–5460, 2019.

[23] J. Barbier and N. Macris, "The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference," *Probability Theory and Related Fields*, vol. 174, pp. 1133–1185, 2017.

[24] F. Guerra and F. L. Toninelli, "The thermodynamic limit in mean field spin glass models," *Communications in Mathematical Physics*, vol. 230, pp. 71–79, 2002.

[25] L. V. Truong, "Fundamental limits and algorithms for sparse linear regression with sublinear sparsity," *J. Mach. Learn. Res.*, vol. 24, pp. 64:1–64:49, 2023.

[26] N. Skantzos, C. Beckmann, and A. Coolen, "Discontinuous recall transitions induced by competition between short- and long-range interactions in recurrent networks," in *Advances in Neural Information Processing Systems*, M. Kearns, S. Solla, and D. Cohn, Eds., vol. 11. MIT Press, 1999.

[27] K. Takeda and Y. Kabashima, "Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals," *EPL (Europhysics Letters)*, vol. 95, no. 1, p. 18006, jun 2011.

[28] ——, "Statistical mechanical analysis of compressed sensing utilizing correlated compression matrix," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 1538–1542.

[29] B. Aubin, B. Loureiro, A. Baker, F. Krzakala, and L. Zdeborov'a, "Exact asymptotics for phase retrieval and compressed sensing with random generative priors," in *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, 2020.

[30] E. W. Tramel, A. Manoel, F. Caltagirone, M. Gabrié, and F. Krzakala, "Inferring sparsity: Compressed sensing using generalized restricted boltzmann machines," *2016 IEEE Information Theory Workshop (ITW)*, pp. 265–269, 2016.

[31] B. Aubin, B. Loureiro, A. Maillard, F. Krzakala, and L. Zdeborová, "The spiked matrix model with generative priors," *IEEE Transactions on Information Theory*, vol. 67, pp. 1156–1181, 2021.

[32] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proceedings of Conference on Information Sciences and Systems*, 2010.

[33] Y. Ma, C. Rush, and D. Baron, "Analysis of approximate message passing with a class of non-separable denoiser and markov random field priors," *IEEE Trans. on Inform. Th.*, vol. 65, no. 11, pp. 7367–7389, 2019.

[34] R. Berthier, A. Montanari, and P. M. Nguyen, "State evolution for approximate message passing with non-separable functions," *Information and Inference: A Journal of IMA*, vol. 9, no. 1, 2020.

[35] D. Donoho, "Compressed sensing," *IEEE Trans. on Inform. Th.*, vol. 52, no. 1, pp. 1289–1306, 2006.

[36] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[37] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. on Inform. Th.*, vol. 62, no. 9, pp. 5117–514, 2016.

[38] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. on Inform. Th.*, vol. 57, no. 2, pp. 764–785, Feb 2011.

[39] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Academy of Sciences (PNAS)*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[40] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *Annals of Mathematical Statistics*, 2020.

[41] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. Academic Press, 1985.

[42] P. Tuominen and R. L. Tweedie, "Markov chains with continuous components," *Proceedings of the London Mathematical Society*, vol. s3-38, no. 1, pp. 18 914–18 919, 1979.

[43] H. Royden and P. Fitzpatrick, *Real Analysis*, 4th ed. Pearson, 2010.

[44] J. Stewart, *Multivariate Calculus*, 6th ed. Cengate Learning, 2008.

[45] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2016, pp. 625–632.

[46] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge Univ. Press, 2010.

[47] L. V. Truong, "Replica analysis of the linear model with markov or hidden markov signal priors," *ArXiv*, vol. abs/2009.13370, 2020.

[48] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley-Interscience, 1995.

[49] S. R. S. Varadhan, "Large deviations," *Annals of Probability*, vol. 36, no. 2, pp. 397–419, 2008.

[50] E. Deutsch and M. Neumann, "Derivatives of the Perron root at an essential nonnegative matrix and the group inverse of an M-matrix," *Journal of Mathematical Analysis and Applications*, vol. 102, pp. 1–29, 1984.

Lan V. Truong (S'12-M'15) received the B.S.E. degree in Electronics and Telecommunications from the Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam, in 2003, and the M.S.E. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, in 2011, and the Ph.D. degree from the Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore, in 2018. He was an Operation and Maintenance Engineer with MobiFone Telecommunications Corporation, Hanoi, for several years. He spent one year as a Research Assistant with the NSF Center for Science of Information and the Department of Computer Science, Purdue University, in 2012. From 2013 to 2015, he was a University Lecturer with the Department of Information Technology Specialization, FPT University, Hanoi, Vietnam. From 2018 to 2019, he was a Research Fellow with the Department of Computer Science, School of Computing, NUS. Since 2020, he has been a Research Associate with the Department of Engineering, University of Cambridge, United Kingdom. His research interests include information theory, machine learning, data science, and probability.