Optimized Markov Chain Monte Carlo for Signal Detection in MIMO Systems: an Analysis of Stationary Distribution and Mixing Time

Babak Hassibi [†], Morten Hansen[*], Alexandros Georgios Dimakis [‡], Haider Ali Jasim Alshamary [§], and Weiyu Xu [§]

Abstract—In this paper we introduce an optimized Markov Chain Monte Carlo (MCMC) technique for solving the integer least-squares (ILS) problems, which include Maximum Likelihood (ML) detection in Multiple-Input Multiple-Output (MIMO) systems. Two factors contribute to the speed of finding the optimal solution by the MCMC detector: the probability of the optimal solution in the stationary distribution, and the mixing time of the MCMC detector. Firstly, we compute the optimal value of the "temperature" parameter, in the sense that the temperature has the desirable property that once the Markov chain has mixed to its stationary distribution, there is polynomially small probability ($1/\mathbf{poly}(N)$, instead of exponentially small) of encountering the optimal solution. This temperature is shown to be at most $O(\sqrt{\mathbf{SNR}}/\ln(N))$ [1], where SNR is the signal-to-noise ratio, and N is the problem dimension. Secondly, we study the mixing time of the underlying Markov chain of the proposed MCMC detector. We find that, the mixing time of MCMC is closely related to whether there is a local minimum in the lattice structures of ILS problems. For some lattices without local minima, the mixing time of the Markov chain is independent of SNR, and grows polynomially in the problem dimension; for lattices with local minima, the mixing time grows unboundedly as SNR grows, when the temperature is set, as in conventional wisdom, to be the standard deviation of noises. Our results suggest that, to ensure fast mixing for a fixed dimension N, the temperature for MCMC should instead be set as $\Omega(\sqrt{\mathbf{SNR}})$ in general. Simulation results show that the optimized MCMC detector efficiently achieves approximately ML detection in MIMO systems having a huge number of transmit and receive dimensions.

I. INTRODUCTION

The problem of performing Maximum Likelihood (ML) decoding in digital communication has gained much attention over the years. These ML decoding problems often reduce to integer least-squares (ILS) problems, which aim to find an integer lattice point closest to received signals. In fact,

[1]In this paper, $\Omega(\cdot)$, $\Theta(\cdot)$, and $O(\cdot)$ are the usual scaling notations as in computer science

the ILS problem is an NP-hard optimization problem appearing in many research areas, for example, communications, global navigation satellite systems, radar imaging, Monte Carlo second-moment estimation, bioinformatics, and lattice design [3], [4]. A computationally efficient way of exactly solving the ILS problem is the sphere decoder (SD) [3], [5]–[8]. It is known that for a moderate problem size and a suitable range of Signal-to-Noise Ratios (SNR), SD has low computational complexity, which can be significantly smaller than an exhaustive search solver. But for a large problem size and fixed SNR, the average computational complexity of SD is still exponential in the problem dimension [9]. So for large problem sizes, (for example massive Multiple-Input Multiple-Output (MIMO) systems with many transmit and receive antennas [10] [11]), SD still has high computational complexity and is thus computationally infeasible. A way to overcome this problem is to use approximate Markov Chain Monte Carlo (MCMC) detectors instead, which can provide the optimal solution asymptotically [12], [13].

Unlike SD, MCMC algorithms perform a random walk over the signal space in the hope of finding the optimal solution. Glauber dynamics is a popular MCMC method which performs the random walk according to the transition probability determined by the stationary distribution of a reversible Markov chain [13] [14]. [15], [16] proposed Glauber dynamics MCMC detectors for data detection in wireless communication (see also the references therein). These MCMC methods are able to provide the optimal solution if they are run for a sufficiently long time; and empirically MCMC methods are observed to provide near-optimal solutions in a reasonable amount of computational time even for large problem dimensions [15], [16]. However, as observed in [16]–[18], unlike sphere decoders , which performs well in high SNR regimes, MCMC detectors often suffer performance degradation in high SNR regimes. In fact, the MCMC detectors in the literature were proposed mostly as practical heuristic detectors for digital communications, and theoretical understandings of the performance and complexity of MCMC detectors for ILS problems are very limited. For example, the mixing time (convergence rate) of the underlying Markov chains of these MCMC detectors, namely how fast these Markov chains mix to its stationary distribution, is not explicitly known. For the MCMC detectors in the literature [15], [16], the conditional

transition probabilities of their underlying Markov chains were directly determined by the posterior likelihood of signal sequences [15], [16]. In other words, the standard deviation of channel noise was naturally applied as the "temperature" of these MCMC detectors [15], [16]. However, it was not clear whether this choice of temperature is optimal, and what effect it will have on the performance and complexity of MCMC detectors.

Two factors contribute to the speed of finding the optimal solution by the MCMC detector: the probability of the optimal solution in the stationary distribution, and the mixing time of the underlying Markov chain for the MCMC detector. In fact, if the optimal solution has a high probability in the stationary distribution, the MCMC detector will very likely encounter the optimal solution when its underlying Markov chain has mixed to its stationary distribution. However, as we will see in this paper, increasing the probability in the stationary distribution of the optimal solution often (even though not always) results in a slow mixing of the underlying Markov chain, namely it takes long time for the Markov chain to reach its stationary distribution. How to balance the mixing time and the stationary distribution for best performance of MCMC detectors is the main subject of this paper.

Our main contributions in this paper are twofold: characterizing the stationary distribution, and bounding the mixing time of MCMC detectors. These results lead to an optimized MCMC detector for solving ILS problems. Firstly, we compute the optimal value of the "temperature" parameter, in the sense that the temperature has the desirable property that once the Markov chain has mixed to its stationary distribution, there is polynomially (and not exponentially) small probability of encountering the optimal solution. This temperature is shown to be at most $O(\sqrt{\text{SNR}}/\ln(N))$, where SNR is the signal-to-noise ratio, and N is the problem dimension. Secondly, we study the mixing time of the underlying Markov chain of the proposed MCMC detector. We find that, the mixing time of MCMC is closely related to whether there is a local minimum in the lattice structures of ILS problems. For some lattices without local minima, the mixing time of the Markov chain is independent of SNR, and grows polynomially in the problem dimension; for lattices with local minima, the mixing time grows unboundedly as SNR grows, when the temperature is set, as in conventional wisdom, to be the standard deviation of noises. We also study the probability that there exist local minima in an ILS problem. For example, the probability of having local minima is $\frac{1}{3} - \frac{1}{\sqrt{5}} + \frac{2\arctan(\sqrt{\frac{5}{3}})}{\sqrt{5}\pi}$ for 2×2 Gaussian MIMO matrices. Simulation results indicate, when the system dimension $N \to \infty$, there seems to be always at least one local minimum with Gaussian MIMO matrices, but we do not have a rigorous proof of this phenomenon. Our theoretical and empirical results suggest that, to ensure fast mixing, for a fixed dimension N, the temperature for MCMC should instead be set as $\Omega(\sqrt{\text{SNR}})$, contrary to conventional wisdom of using the standard deviation of channel noises [15], [16]. Our simulation results show that the optimized MCMC

detector efficiently achieves approximately ML detection in MIMO systems having a huge number of transmit and receive dimensions.

We caution, however, that we have not been able to prove the scaling of the mixing time in terms of system dimension N for integer least-squares problems. The question whether MCMC detectors mix in polynomial time over N remains an open problem.

The paper is organized as follows. In Section II we present the system model that will be used throughout the paper. The MCMC methods and background knowledge on Markov chain mixing time are described in Section III. In Section IV we analyze the probability of error for the ML detector. Section V treats the optimal selection of the temperature parameter α. Section VI, VII,VIII and IX derive bounds on the mixing time and discuss how to optimize MCMC parameters to ensure fast mixing. Simulation results are given in Section X.

II. SYSTEM MODEL

We consider a real-valued block-fading MIMO antenna system, with N transmit and N receive dimensions, with know channel coefficients. The received signal $\mathbf{y} \in \mathbb{R}^N$ can be expressed as

$$\mathbf{y} = \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x} + \boldsymbol{v} \ , \tag{1}$$

where $\mathbf{x} \in \Xi^N$ is the transmitted signal, and Ξ denotes the constellation set. To simplify the derivations in the paper we will assume that $\Xi = \{\pm 1\}$. $\boldsymbol{v} \in \mathbb{R}^N$ is the noise vector where each entry is Gaussian $\mathcal{N}(0,1)$ and independent identically distributed (i.i.d.), and $\mathbf{H} \in \mathbb{R}^{N \times N}$ denotes the channel matrix with i.i.d. $\mathcal{N}(0,1)$ entries. (In general, $\mathbf{H}$ can be any matrix, however, for analysis purposes we will focus on $\mathbf{H}$ with i.i.d. Gaussian elements.) The signal-to-noise ratio is defined as

$$\text{SNR} = \frac{\mathcal{E}\left\|\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x}\right\|^2}{\mathcal{E}\|\boldsymbol{v}\|^2} \ , \tag{2}$$

which is done in order to take into account the total transmit energy. For analysis purposes we will focus on the regime where $\text{SNR} > 2\ln(N)$, in order to get the probability of error of the ML detector to go to zero. Without loss of generality, we will assume that the all-minus-one vector was transmitted, $\mathbf{x} = -\mathbf{1}$. Therefore

$$\mathbf{y} = \boldsymbol{v} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{1} \ . \tag{3}$$

We are considering a minimization of the average error probability $P(\mathbf{e}) \triangleq P(\hat{\mathbf{x}} \neq \mathbf{x})$, which is obtained by performing Maximum Likelihood Sequence Detection (here simply referred to as ML detection) given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \Xi^N} \left\|\mathbf{y} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x}\right\|^2 \ . \tag{4}$$

III. MCMC Detector

One way of solving the optimization problem given in (4) is by using Markov Chain Monte Carlo (MCMC) detectors, which asymptotically converge to the optimal solution if the detector follows a reversible Markov chain [19]. We first describe our proposed MCMC detector based on reversible Markov chain, and then compare it with existing MCMC detectors in the literature.

A. Reversible MCMC Detector

In this paper, we mainly focus on a MCMC detector which follows a reversible Markov chain and asymptotically converges to the stationary distribution [19]. Under the stationary distribution, the MCMC detector has a certain probability of visiting the optimal solution. So if run for a sufficiently long time, the MCMC detector will be able to find the optimal solution to (4).

For this MIMO detection problem (4), the MCMC detector starts with a certain N-dimensional feasible vector $\hat{\mathbf{x}}^{(0)}$ among the set $\{-1, +1\}^N$ of cardinality 2^N. Then the MCMC detector performs a random walk over $\{-1, +1\}^N$ based on the following reversible Markov chain. Assume that we are at time index l and the current state of the Markov chain is $\hat{\mathbf{x}}^{(l)} \in \{-1, +1\}^N$. In the next step, the Markov chain picks one random position index j uniformly out of $\{1, 2, ..., N\}$, and keeps the symbols of $\hat{\mathbf{x}}^{(l)}$ at other positions fixed. Then the MCMC detector computes the conditional probability of transferring to each constellation point at the j-th index. With the symbols at the $(N-1)$ other positions fixed, the probability that the j-th symbol adopts the value ω, is given by

$$
p\left(\hat{\mathbf{x}}_j^{(l+1)} = \omega \,|\, \theta\right) = \frac{e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\hat{\mathbf{x}}_{j|\omega}\right\|^2}}{\sum\limits_{\hat{\mathbf{x}}_{j|\bar\omega} \in \Xi} e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\hat{\mathbf{x}}_{j|\bar\omega}\right\|^2}}\ ,\quad (5)
$$

where $\hat{\mathbf{x}}_{j|\omega}^T \triangleq \left[\hat{\mathbf{x}}_{1:j-1}^{(l)}, \omega, \hat{\mathbf{x}}_{j+1:N}^{(l)}\right]^T$ and $\theta = \left\{\hat{\mathbf{x}}^{(l)}, j, \mathbf{y}, \mathbf{H}\right\}$. So conditioned on the j-th position is chosen, the MCMC detector will with probability $p\left(\hat{\mathbf{x}}_j^{(l+1)} = \omega \,|\, \theta\right)$ transition to ω at the j-th position index. The initialization of the symbol vector $\hat{\mathbf{x}}^{(0)}$ can either be chosen randomly or as other heuristic solutions.

Our algorithm is summarized as follows in Algorithm 5.

For this type of MCMC detector, what one cares about is the probability that such an algorithm encounters the true transmitted signal within a certain number of iterations. In general, determining this probability within a certain number of iterations is difficult. However, things are relatively easy when we assume that the underlying Markov chain has mixed to the steady state distribution, which is easy to write down, and therefore in steady state it is easy to determine this probability P_{en} of encountering the true transmitted signal. Therefore, an upper bound on the expected time to find the optimal solution is determined by the mixing time (the time it takes to to reach the steady state) of the underlying Markov

---

Algorithm 1: MCMC detector based on reversible Markov chain

Input: $\mathbf{y}$, $\mathbf{H}$, initialization vector $\hat{\mathbf{x}}^{(0)}$, decision vector $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(0)}$ and the number of iterations n

1 **for** $i = 1$ **to** n **do**
2   pick a uniformly random position index j out of $\{1, 2, ..., N\}$
3   keep the symbols of $\hat{\mathbf{x}}^{(i-1)}$ at the $(N-1)$ other positions fixed, transition the j-th symbol of $\hat{\mathbf{x}}^{(i-1)}$ to ω with probability $p\left(\hat{\mathbf{x}}_j^{(l+1)} = \omega \,|\, \theta\right)$ specified in (5), for every $\omega \in \{-1, +1\}$
4   denote the new vector by $\hat{\mathbf{x}}^{(i)}$
5   if $\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{(i)}\|_2^2 < \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_2^2$, update $\hat{\mathbf{x}} := \hat{\mathbf{x}}^{(i)}$

---

chain, and the inverse of the probability P_{en} of encountering the true transmitted signal in the steady state.

We remark that α represents a tunable positive parameter which controls the mixing time of the Markov chain, and this parameter is also sometimes called the "temperature". If we let $\alpha \to \infty$, the MCMC detector is a just a uniform random walk in the signal space, namely in each iteration the detector choose constellation points with equal probabilities, and the underlying Markov chain quickly mixes to its steady state [14]. When α is close to 0, the MCMC detector will eventually "reside" at the optimal solution, but it may take a very long time to get there from an initial suboptimal signal vector.

On the one hand, the smaller α is, the larger the stationary probability for the optimal solution will be, and the easier it is for the MCMC detector to find the optimal solution in the stationary distribution. On the other hand, as α gets smaller, it often takes a long time for the Markov chain to converge to its stationary distribution. In fact, as we will show in the paper, there is often a lower bound on α, in order to ensure the fast mixing of the Markov chain to its stationary distribution.

B. Comparisons with conventional MCMC detectors

Our proposed MCMC detector is different from conventional MCMC detectors [15], [16]. In [15], [16], the conditional transition probabilities of the underlying Markov chains were directly determined by the posterior likelihood of data sequences. In other words, the "temperature" α of these MCMC detectors is directly set as the standard deviation of channel noise [15], [16]. In this paper, however, we have the freedom of optimizing this temperature parameter α.

Our proposed method is also very different from simulated annealing techniques where the temperature is slowly reduced until the detector converges to an acceptable solution. In our MCMC detector, the temperature is set as a *fixed* value, and we care about a fast mixing of the underlying Markov chain to a stationary probability distribution and a big enough probability of encountering the transmitted signal in steady state.

C. Mixing time

It is not hard to see that the Markov chain of MCMC detector is reversible and has 2^N states with the stationary distribution $e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y}-\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\hat{\mathbf{x}}\right\|^2}$ (without normalization) for a state $\hat{\mathbf{x}}$. The $2^N \times 2^N$ transition matrix is denoted by P, and the element $P_{i,j}$ in the i-th ($1 \le i \le N$) row and j-th ($1 \le j \le N$) column is the probability of transferring to state j conditioned on the previous state is i. So each row of P sums up to 1 and the transition matrix after t iterations is P^t. We denote the vector for the stationary distribution as π. Then for an $\epsilon > 0$, the mixing time $t(\epsilon)$ is a parameter describing how long it takes for the Markov chain to get close to the stationary distribution [14], namely,

$$t_{mix}(\epsilon) := \min\{t : \max_{\tilde{\mathbf{x}}} \|P^t(\tilde{\mathbf{x}}, \cdot) - \pi\|_{TV} \le \epsilon\},$$

where $\|\mu - \nu\|_{TV}$ is the usual total variation distance between two distributions μ and ν over the state space $\{+1, -1\}^N$.

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{\mathbf{z} \in \{+1, -1\}^N} |\mu(\mathbf{z}) - \nu(\mathbf{z})|.$$

The mixing time is closely related to the spectrum of the transition matrix P. More precisely, for a reversible Markov chain, its mixing time is generally small when the gap between the largest and the second largest eigenvalue of P, namely $1 - \lambda_2$, is large. The inverse of this gap, $\frac{1}{1-\lambda_2}$, is called the relaxation time for this Markov chain.

D. Sequential Markov Chain Monte Carlo Detectors

In this paper, for simplicity of implementations, we also consider a sequential MCMC detector, especially in numerical simulations. The only difference between sequential MCMC detectors and reversible MCMC detectors is the way they choose the position index to update. Sequential MCMC detectors can have many block iterations. We define one *block iteration* of the sequential MCMC detector as an sequential update of all the N indices $\{1, \ldots, N\}$ in the estimated symbol vector $\hat{\mathbf{x}}$, starting from $j = 1$ to $j = N$. Namely, in one block iteration, we update N indices. For each index j, the updating rule for the sequential MCMC detector is the same as the reversible MCMC detector. We remark, however, that the mixing time results in this paper are only for reversible MCMC detectors.

E. Complexity of the MCMC detector

The conditional probability for the j-th symbol in (5) can be computed efficiently by reusing the result obtained in earlier iterations, when we evaluate $\left\|\mathbf{y} - \sqrt{\text{SNR}/N}\mathbf{H}\hat{\mathbf{x}}_{j|\omega}\right\|^2$. Since we are only changing the j-th symbol in the symbol vector, the difference $\mathbf{d}^{(l)} \triangleq \mathbf{y} - \sqrt{\text{SNR}/N}\mathbf{H}\hat{\mathbf{x}}_{j|\omega}$ can be expressed as

$$\mathbf{d}^{(l)} = \mathbf{d}^{(l-1)} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{h}_j \Delta s_{j|\omega} \ , \tag{6}$$

where l is the index for the number of iterations, $\Delta x_{j|\omega} \triangleq x_{j|\omega}^{(l)} - x_{j|\bar{\omega}}^{(l-1)}$, and $\mathbf{h}_j$ is the j-th column of $\mathbf{H}$. Thus, the computation of conditional probability when changing the symbol in the j-th position costs $2N$ operations[2], where we define an operation as a Multiply and Accumulate (MAC) instruction. This leads to a complexity of $O\left(2N[|\Xi|-1]\right)$ operations per iteration.

F. MCMC sampling using QR- or QL-factorization

In the case where the number of iterations in the MCMC detector is sufficiently larger than the system size, the complexity of MCMC detector can be reduced even further using a QR- or QL-factorization of the channel matrix, $\mathbf{H} = \tilde{\mathbf{Q}}\mathbf{R} = \mathbf{QL}$, such that the optimization problem in (4) becomes

$$\min_{s \in \Xi^N} \ \left\|\tilde{\mathbf{y}} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{Lx}\right\|^2 \ , \tag{7}$$

where $\hat{\mathbf{y}} \triangleq \mathbf{Q}^T\mathbf{y}$. Since $\mathbf{L}$ is a lower triangular matrix, the product $\mathbf{Lx}$ requires less operations compared to a full channel matrix. Suppose that we need to update position index j at the current iteration and assume $\mathbf{d}^{(l-1)}$ is known, we only need to compute the indices from j to N in $\mathbf{d}^{(l)}$, since these are the only non-zero elements in $\mathbf{L}_{1:N,j}\Delta x_{j|\Xi}$. Thus, for a square channel matrix of size N the complexity of one iteration in the MCMC detector can roughly be reduced to half the number of operations, namely $O\left(N[|\Xi|-1]\right)$. This computation saving should be compared with the complexity of performing the QL-factorization, which requires $O\left(N^3\right)$, and therefore, in cases where the number of iterations is $k > N^2/\left(|\Xi|-1\right)$, we can achieve a complexity reduction.

G. Norm-2 MCMC Sampler

In this paper, we also propose a new MCMC detector called norm-2 MCMC detector. For this new MCMC detector, we can more easily lower bound its mixing time. We remark, however, in most parts of this paper, "the MCMC detector" refers to the squared-norm-2 MCMC detector in Subsection III-A.

The norm-2 MCMC detector is mostly identical to the squared-norm-2 MCMC detector, except for the computation of transition probability in (5). Instead of using (5), the transition probability for norm-2 MCMC detector is given by

$$p\left(\hat{\mathbf{x}}_j^{(l+1)} = \omega \,|\theta\right) = \frac{e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y}-\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\hat{\mathbf{x}}_{j|\omega}\right\|_2}}{\sum\limits_{\hat{\mathbf{x}}_{j|\bar{\omega}} \in \Xi} e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y}-\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\hat{\mathbf{x}}_{j|\bar{\omega}}\right\|_2}} \ , \tag{8}$$

where $\tilde{\mathbf{x}}_{j|\omega}^T \triangleq \left[\hat{\mathbf{x}}_{1:j-1}^{(l)}, \omega, \hat{\mathbf{x}}_{j+1:N}^{(l)}\right]^T$ and $\theta = \left\{\hat{\mathbf{x}}^{(l)}, j, \mathbf{y}, \mathbf{H}\right\}$. So we only use the ℓ_2 norm in the exponent of the transition probability.

---

[2]We need to compute both the product $\mathbf{h}_j \Delta x_{j|\omega}$ and the inner product $(\mathbf{d}^{(l)})^T\mathbf{d}^{(l)}$.

IV. Probability of Error

First, we would like to derive the probability of error for ML detection in MIMO systems, and then use the results to characterize the SNR regime of interest. The error probability is calculated by averaging over the random matrices H and random noises. Before we derive the probability of error for the ML detector, we will state a lemma which we will make repeated use of.

Lemma IV.1 (Gaussian Integral). *Let* $\mathbf{v}$ *and* $\mathbf{x}$ *be independent Gaussian random vectors with distribution* $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ *each. Then*

$$\mathcal{E}\left\{e^{\eta\left(\|\mathbf{v}+a\mathbf{x}\|^2 - \|\mathbf{v}\|^2\right)}\right\} = \left(\frac{1}{1 - 2a^2\eta(1 + 2\eta)}\right)^{N/2}. \quad (9)$$

Proof: See Appendix XI-A for a detailed proof. ∎

Let us first look at the probability of error using maximum likelihood detection. We will make an error if there exists a vector $\mathbf{x} \neq -\mathbf{1}$ such that

$$\left\|\mathbf{y} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x}\right\|^2 \leq \left\|\mathbf{y} + \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{1}\right\|^2 = \|\boldsymbol{v}\|^2.$$

In other words,

$$
\begin{aligned}
P_e &= \text{Prob}\left(\left\|\mathbf{y} - \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x}\right\|^2 \leq \|\boldsymbol{v}\|^2\right) \\
&= \text{Prob}\left(\left\|\boldsymbol{v} + \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}(-\mathbf{1} - \mathbf{x})\right\|^2 \leq \|\boldsymbol{v}\|^2\right),
\end{aligned}
$$

for some $\mathbf{x} \neq -\mathbf{1}$, which can be formulated as

$$P_e = \text{Prob}\left(\left\|\boldsymbol{v} + 2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2 \leq \|\boldsymbol{v}\|^2\right),$$

for some $\boldsymbol{\delta} \neq 0$, where $\boldsymbol{\delta} \triangleq \frac{1}{2}(-\mathbf{1} - \mathbf{x})$. Note that in the above equation $\boldsymbol{\delta}$ is a vector of zeros and -1's. Now using the union bound

$$P_e \leq \sum_{\boldsymbol{\delta} \neq 0} \text{Prob}\left(\left\|\boldsymbol{v} + 2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2 \leq \|\boldsymbol{v}\|^2\right). \quad (10)$$

We will use the Chernoff bound to bound the quantity inside the summation. Thus,

$$
\begin{aligned}
&\text{Prob}\left(\left\|\boldsymbol{v} + 2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2 \leq \|\boldsymbol{v}\|^2\right) &(11a) \\
&\leq \mathcal{E}\left\{e^{-\beta\left(\left\|\boldsymbol{v}+2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2 - \|\boldsymbol{v}\|^2\right)}\right\} &(11b) \\
&= \left(\frac{1}{1 + 8\frac{\text{SNR}\|\boldsymbol{\delta}\|^2}{N}\beta(1 - 2\beta)}\right)^{N/2}, &(11c)
\end{aligned}
$$

where $\beta \geq 0$ is the Chernoff parameter, and where we have used Lemma IV.1 with $\eta = -\beta$ and $a = 2\sqrt{\frac{\text{SNR}\|\boldsymbol{\delta}\|^2}{N}}$, since

$$\mathcal{E}\left\{\left(2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right)\left(2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right)^T\right\} = 4\frac{\text{SNR}\|\boldsymbol{\delta}\|^2}{N}\mathbf{I}_N.$$

The optimal value for β is $\frac{1}{4}$, which yields the tightest bound

$$\text{Prob}\left(\left\|\boldsymbol{v} + 2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2 \leq \|\boldsymbol{v}\|^2\right) \leq \left(\frac{1}{1 + \frac{\text{SNR}\|\boldsymbol{\delta}\|^2}{N}}\right)^{N/2}. \quad (12)$$

Note that this depends only on $\|\boldsymbol{\delta}\|^2$, the number of nonzero entries in $\boldsymbol{\delta}$. Plugging this into the union bound yields

$$P_e \leq \sum_{i=1}^{N} \binom{N}{i}\left(\frac{1}{1 + \frac{\text{SNR}i}{N}}\right)^{N/2}. \quad (13)$$

Let us first look at the linear (i.e., i proportional to N) terms in the above sum. Thus,

$$\binom{N}{i}\left(\frac{1}{1 + \frac{\text{SNR}i}{N}}\right)^{N/2} \approx e^{NH(\frac{i}{N}) - \frac{N}{2}\ln\left(1 + \frac{\text{SNR}i}{N}\right)},$$

where $H(\cdot)$ is entropy in "nats". Clearly, if

$$\lim_{N\to\infty}\text{SNR} = \infty,$$

then the linear terms go to zero (superexponentially fast).

Let us now look at the sublinear terms. In particular, let us look at $i = 1$:

$$N\left(\frac{1}{1 + \frac{\text{SNR}}{N}}\right)^{N/2} \approx Ne^{-\text{SNR}/2}.$$

Clearly, to have this term go to zero, we require that SNR $> 2\ln N$.

A similar argument shows that all other sublinear terms also go to zero, and so we have:

Lemma IV.2 (SNR scaling). *If SNR $> 2\ln N + f(N)$, where $f(N)$ is an arbitrary function that goes to ∞ as $N \to \infty$, then $P_e \to 0$ as $N \to \infty$.*

V. Computing the Optimal α

In this section, we derive the optimal value of the "temperature" parameter which controls the mixing time of the underlying Markov chain. The temperature has the desirable property that once the Markov chain has mixed to steady state, there is only polynomially (and not exponentially) small probability of encountering the optimal solution.

A. Mean of π_{-1}

In the following section we compute the expected value of the stationary probabilities of the states, where the expectation is taken over random Gaussian H and noises. More specifically, we are examining the probability of state $\mathbf{x} = -\mathbf{1}$, denoted by π_{-1} (recall that we assumed that $-\mathbf{1}$ is transmitted symbol vector).

This calculation has a lot in common with the one given in Section IV. Note that

$$\pi_{-1} = \frac{e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y}+\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{1}\right\|^2}}{\sum_{\mathbf{x}} e^{-\frac{1}{2\alpha^2}\left\|\mathbf{y}+\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x}\right\|^2}} \tag{14a}$$

$$= \frac{e^{-\frac{1}{2\alpha^2}\|\boldsymbol{v}\|^2}}{\sum_{\mathbf{x}} e^{-\frac{1}{2\alpha^2}\left\|\boldsymbol{v}+\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}(\mathbf{x}-\mathbf{1})\right\|^2}} \tag{14b}$$

$$= \frac{e^{-\frac{1}{2\alpha^2}\|\boldsymbol{v}\|^2}}{\sum_{\boldsymbol{\delta}} e^{-\frac{1}{2\alpha^2}\left\|\boldsymbol{v}+2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2}} \tag{14c}$$

$$= \frac{1}{\sum_{\boldsymbol{\delta}} e^{-\frac{1}{2\alpha^2}\left(\left\|\boldsymbol{v}+2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2-\|\boldsymbol{v}\|^2\right)}} \;, \tag{14d}$$

where $\boldsymbol{\delta}$ is a vector of zeros and ones.

Now, using Jensen's inequality and the convexity of $\frac{1}{t}$ when $t > 0$,

$$\mathcal{E}\{\pi_{-1}\} \geq \frac{1}{\mathcal{E}\left\{\frac{1}{\pi_{-1}}\right\}} \tag{15a}$$

$$= \frac{1}{\mathcal{E}\left\{\sum_{\boldsymbol{\delta}} e^{-\frac{1}{2\alpha^2}\left(\left\|\boldsymbol{v}+2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2-\|\boldsymbol{v}\|^2\right)}\right\}} \tag{15b}$$

$$= \frac{1}{\sum_{\boldsymbol{\delta}} \mathcal{E}\left\{e^{-\frac{1}{2\alpha^2}\left(\left\|\boldsymbol{v}+2\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\boldsymbol{\delta}\right\|^2-\|\boldsymbol{v}\|^2\right)}\right\}} \tag{15c}$$

$$= \frac{1}{1 + \sum_{\boldsymbol{\delta}\neq 0}\left(\frac{1}{1+4\frac{\text{SNR}\|\boldsymbol{\delta}\|^2}{N}\frac{1}{\alpha^2}(1-\frac{1}{\alpha^2})}\right)^{N/2}} \tag{15d}$$

$$= \frac{1}{1 + \sum_{i=1}^{N}\binom{N}{i}\left(\frac{1}{1+\frac{\beta i}{N}}\right)^{N/2}} \;. \tag{15e}$$

In (15d) we have used Lemma IV.1 and in (15e) we have defined $\beta \triangleq 4\text{SNR}\frac{1}{\alpha^2}(1-\frac{1}{\alpha^2})$. While it is possible to focus on the linear and sublinear terms in the above summation separately, to give conditions for $\mathcal{E}\{\pi_{-1}\}$ to have the form of $1/\text{poly}(N)$, we will be interested in the exact exponent of the poly and so we need a more accurate estimate. To do this we shall use saddle point integration. To this end, note that

$$\binom{N}{i}\left(\frac{1}{1+\frac{\beta i}{N}}\right)^{N/2} = e^{NH(\frac{i}{N})-\frac{N}{2}\ln\left(1+\frac{\beta i}{N}\right)} \;,$$

again $H(\cdot)$ represents the entropy in "nats". And so the summation in the denominator of (15e) can be approximated as a Stieltjes integral:

$$\sum_{i=1}^{N}\binom{N}{i}\left(\frac{1}{1+\frac{\beta i}{N}}\right)^{N/2} = N\sum_{i=1}^{N} e^{NH(\frac{i}{N})-\frac{N}{2}\ln\left(1+\frac{\beta i}{N}\right)}\frac{1}{N} \tag{16a}$$

$$= N\int_0^1 e^{NH(x)-\frac{N}{2}\ln(1+\beta x)}dx \;. \tag{16b}$$

For large N, this is a saddle point integral and can be approximated by the formula

$$\int_0^1 e^{Nf(x)}dx \approx \sqrt{\frac{2\pi}{N|f''(x_0)|}}e^{Nf(x_0)} \;, \tag{17}$$

where x_0 is the saddle point of $f(\cdot)$, i.e., $f'(x_0) = 0$. In our case,

$$f(x) = -x\ln x - (1-x)\ln(1-x) - \frac{1}{2}\ln(1+\beta x) \;,$$

and so

$$f'(x) = \ln\frac{1-x}{x} - \frac{1}{2}\frac{\beta}{1+\beta x} \;.$$

In general, it is not possible to solve for $f'(x_0) = 0$ in closed form. However, in our case, we assume that $\beta = 4\text{SNR}\frac{1}{\alpha^2}(1-\frac{1}{\alpha^2}) \gg 1$ (In fact, we must have $\beta \to \infty$ as $N \to \infty$. Otherwise, (16) will be exponential in N). In this case, it is not too hard to verify that the saddle point is given by

$$x_0 \approx e^{-\frac{\beta}{2}} \;. \tag{18}$$

And hence

$$f(x_0) = -e^{-\frac{\beta}{2}}\ln e^{-\frac{\beta}{2}} - (1-e^{-\frac{\beta}{2}})\ln(1-e^{-\frac{\beta}{2}})$$
$$\quad - \frac{1}{2}\ln(1+\beta e^{-\frac{\beta}{2}})$$
$$\approx \frac{\beta}{2}e^{-\frac{\beta}{2}} + e^{-\frac{\beta}{2}} - \frac{1}{2}\beta e^{-\frac{\beta}{2}}$$
$$= e^{-\frac{\beta}{2}} \;,$$

and further plugging x_0 into

$$f''(x) = -\frac{1}{x} - \frac{1}{1-x} - \frac{1}{2}\frac{\beta^2}{(1+\beta x)^2} \;,$$

yields

$$f''(x_0) \approx -e^{\frac{\beta}{2}} - 1 + \frac{1}{2}\beta^2 \approx -e^{\frac{\beta}{2}} \;. \tag{19}$$

Replacing these into the saddle point expression in (17) show that

$$\sum_{i=1}^{N}\binom{N}{i}\left(\frac{1}{1+\frac{\beta i}{N}}\right)^{N/2} \approx \sqrt{2\pi/N}\exp\left(Ne^{-\frac{\beta}{2}} - \frac{\beta}{4}\right) \;. \tag{20}$$

We want $\mathcal{E}\{\pi_{-1}\}$ to behave as $\frac{1}{N^\zeta}$ and according to (14) this means that we want the expression in (20) to behave as N^ζ, where ζ is a positive number. Let us take

$$e^{Ne^{-\frac{\beta}{2}}} = N^\zeta \;.$$

Solving for β yields

$$\beta = 2(\ln N - \ln \ln N - \ln \zeta). \qquad (21)$$

Incidentally, this choice of β yields

$$e^{-\frac{\beta}{4}} \approx \frac{1}{\sqrt{N}}. \qquad (22)$$

Finally, this choice of β means that we have

$$4\text{SNR}\frac{1}{\alpha^2}\left(1 - \frac{1}{\alpha^2}\right) = 2(\ln N - \ln \ln N - \ln \zeta) \, ,$$

and so we have the following result.

Lemma V.1 (Mean of π_{-1}). *As $N \to \infty$, if α is chosen such that*

$$\frac{\alpha^2}{1 - \frac{1}{\alpha^2}} = \frac{4SNR}{2(\ln N - \ln \ln N - \ln \zeta)} \, , \qquad (23)$$

then

$$\mathcal{E}\left\{\frac{1}{\pi_{-1}}\right\} \le N^\zeta \, . \qquad (24)$$

and

$$\mathcal{E}\left\{\pi_{-1}\right\} \ge N^{-\zeta} \, . \qquad (25)$$

When we have an upper bound on $\mathcal{E}\left\{\frac{1}{\pi_{-1}}\right\}$, we can then use the Markov inequality to give upper bounds on the probability that $\frac{1}{\pi_{-1}}$ exceeds a certain threshold. More precisely, we have $P(\frac{1}{\pi_{-1}} > N^{\gamma'}) \le \mathcal{E}\left\{\frac{1}{\pi_{-1}}\right\}/N^{\gamma'} \le N^{-(\gamma'-\zeta)}$ for any γ'. This means that with probability close to 1 as $N \to \infty$, the expected time to encounter the transmitted signal in steady state is no bigger than $N^{\gamma'}$, for every $\gamma' > \zeta$.

B. Value of α

In this subsection we investigate how α behaves as a function of the SNR and the system dimension, if α is chosen according to (23). In general, the larger α is, the faster the Markov chain mixes. However, choosing α any larger than this means that the probability of finding the optimal solution in stationary distribution is exponentially small. Thus, when choosing the value of α, there is a trade-off between faster mixing time of the Markov chain (due to an increase of α), and faster encountering the optimal solution in stationary distribution. In the following, we evaluate (23) with $\zeta = i$, denoted as $\alpha_{\zeta=i}$ and we also approximate α in (23) by neglecting the terms $\ln \ln(N)$ and $\ln(\zeta)$, leading to

$$\frac{\tilde{\alpha}^4}{\tilde{\alpha}^2 - 1} = \frac{2\text{SNR}}{\ln(N)} \, . \qquad (26)$$

From (26) we see that

$$\alpha^2 = \frac{\text{SNR}}{\ln(N)} \pm \sqrt{\left(\frac{\text{SNR}}{\ln(N)}\right)^2 - 2\frac{\text{SNR}}{\ln(N)}} \, , \qquad (27)$$

which implies that $\tilde{\alpha}$ becomes complex when $\text{SNR} < 2\ln(N)$. However, as stated in Section II we focus on $\text{SNR} > 2\ln(N)$. Since we are solving a quadratic equation we get two values of α^2, representing the region in which (25) is satisfied. Based on the considerations given above, we prefer the value of α^2
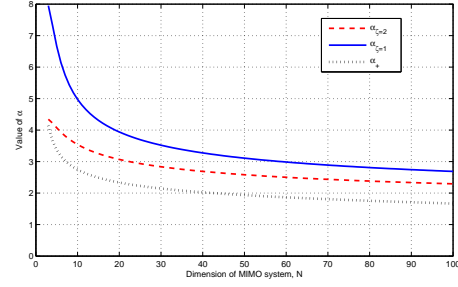

Figure 1: Value of α vs. system size N, for SNR = 10 dB.

obtained by the plus sign in (27), denoted α_+^2, in order to achieve the fastest mixing time. In Figure 1 the values of $\alpha_{\zeta=2}$, $\alpha_{\zeta=1}$, and α_+ have been plotted as a function of the system dimension for SNR = 10 dB. Our simulations also suggest that the computed value of α is very close to the optimal choice, even in the case where the condition $\text{SNR} > 2\ln(N)$ is not satisfied.

C. Mixing time of Markov Chain

So far, we have examined the largest possible α such that the optimal sequence has a reasonable stationary probability. However, all this was based on assuming that we have reached the stationary distribution. As α also affects the speed of getting to the stationary distribution, it is interesting to quantify the mixing time of the Markov chain of MCMC detectors. In the next sections, we will discuss how the mixing time is related to α and the underlying lattice structures in ILS problems.

VI. MIXING TIME WITH ORTHOGONAL MATRICES

Starting from this section, we consider the mixing time for MCMC for ILS problems and study how the mixing time for ILS problem depends on the linear matrix structure and SNR. As a first step, we consider a linear matrix $\mathbf{H}$ with orthogonal columns. As shown later, the mixing time for this matrix has an upper bound independent of SNR. In fact, this is a general phenomenon for ILS problems without local minima.

For simplicity, we use $\acute{\mathbf{H}}$ to represent $\sqrt{\frac{\text{SNR}}{N}}\mathbf{H}$, and the model we are currently considering is

$$\mathbf{y} = \acute{\mathbf{H}}\mathbf{x} + \mathbf{v}. \qquad (28)$$

When the SNR increases, we simply increase the amplitude of elements in $\acute{\mathbf{H}}$. We will also incorporate the SNR term into $\acute{\mathbf{H}}$ this way in the following sections unless stated otherwise.

Theorem VI.1. *Independent of the temperature α and SNR, the mixing time of the MCMC detector for orthogonal-column ILS problems is upper bounded by $N\log(N) + \log(1/\epsilon)N$.*

This theorem is an extension of the mixing time for regular random walks on an N-dimensional hypercube [14]. The only difference here is that the transition probability follows (5) and that the transition probability depends on SNR. Under

orthogonal columns, the ILS problem has no local minimum, since $\mathbf{H}^T\mathbf{H}$ is a diagonal matrix in the expansion of $\|\mathbf{y}-\acute{\mathbf{H}}\mathbf{x}\|_2^2$.

Proof: When the j-th index was selected for updating in the MCMC detector, since the columns of $\acute{\mathbf{H}}$ are orthogonal to each other, the probability of updating $\mathbf{x}_j$ to -1 is $\frac{1}{1+e^{\frac{2\mathbf{y}^T\mathbf{h}_j}{\alpha^2}}}$. We note that this probability is independent of the current state of Markov chain $\hat{\mathbf{x}}$. So we can use the classical coupling idea to get an upper bound on the mixing time of this Markov chain.

Consider two separate Markov chains starting at two different states $\mathbf{x}_1$ and $\mathbf{x}_2$. These two chains follow the same update rule according to (5). By using the same random source, in each step they select the same position index for updating, and they update that position to the same symbol. Let τ_{couple} be the first time the two chains come to the same state. Then by a classical result, the total variation distance

$$d(t) = \max_{\tilde{\mathbf{x}}} \|P^t(\tilde{\mathbf{x}},\cdot)-\pi\|_{TV} \leq \max_{\mathbf{x}_1,\mathbf{x}_2} p_{\mathbf{x}_1,\mathbf{x}2}\{\tau_{couple} > t\}. \quad (29)$$

Note that the coupling time is just time for collecting all of the positions where $\mathbf{x}_1$ and $\mathbf{x}_2$ differ, as in the coupon collector problem. From the famous coupon collector problem [14], for any $\mathbf{x}_1$ and $\mathbf{x}_2$,

$$d(N\log(N)+cN) \leq p_{\mathbf{x}_1,\mathbf{x}_2}\{\tau_{couple} > N\log(N)+cN\} \leq e^{-c}. \quad (30)$$

So the conclusion follows. ∎

VII. MIXING TIME WITH LOCAL MINIMA

In this section, we consider the mixing time for ILS problems which have local minima besides the global minimum point. Our main results are that local minima greatly affect the mixing time of MCMC detectors, and rigorous statements are given in Theorem VII.4. First, we give the definition of a local minimum.

Definition VII.1. *A local minimum $\tilde{\mathbf{x}}$ is a state such that $\tilde{\mathbf{x}}$ is not a global minimizer for $\min_{\mathbf{s}\epsilon\{-1,+1\}^N} \|\mathbf{y} - \acute{\mathbf{H}}\mathbf{s}\|^2$; and any of its neighbors which differ from $\tilde{\mathbf{x}}$ in only one position index, denoted by $\tilde{\mathbf{x}}'$, satisfies $\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2 > \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2$.*

We will use the following theorem about the spectral gap of Markov chain to evaluate the mixing time.

Theorem VII.2 (Jerrum and Sinclair (1989) [20], Lawler and Sokal (1988) [21], [14])**.** *Let λ_2 be the second largest eigenvalue of a reversible transition matrix P, and let $\gamma = 1 - \lambda_2$. Then*

$$\frac{\Phi_*^2}{2} \leq \gamma \leq 2\Phi_*,$$

where Φ_* is the bottleneck ratio (also called conductance, Cheeger constant, and isoperimetric constant) defined as

$$\Phi_* = \min_{\pi(S)\leq\frac{1}{2}} \frac{Q(S,S^c)}{\pi(S)}.$$

Here S is any subset of the state spaces with stationary measure no bigger than $\frac{1}{2}$, S^c is its complement set, and $Q(S,S^c)$ is the probability of moving from S to S^c in one step when starting with the stationary distribution.

Theorem VII.3. *If there is a local minimum $\tilde{\mathbf{x}}$ in an integer least-squares problem and we denote its neighbor differing only at the j-th $(1 \leq j \leq N)$ location as $\tilde{\mathbf{x}}_j$, then the mixing time of the MCMC detector is at least*

$$t_{mix}(\epsilon) \geq \log(\frac{1}{2\epsilon})(\frac{1}{\gamma} - 1), \quad (31)$$

where

$$\gamma = \sum_{j=1}^{N} \frac{2}{N} \frac{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}} \quad (32)$$

The parameter γ is upper bounded by

$$\frac{2}{1 + e^{\frac{\min_j \|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2-\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}} \quad (33)$$

Proof: We apply Theorem VII.2 to prove this result. We take a local minimum point $\tilde{\mathbf{x}}$ as the single element in the bottle-neck set S. Since $\tilde{\mathbf{x}}$ is a local minimum, $\pi(S) \leq \frac{1}{2}$.

$$Q(S,S^c) = \frac{\pi(S)}{N} \sum_{j=1}^{N} \frac{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}} \quad (34)$$

Dividing by $\pi(S)$, by the definition of Φ_*

$$\Phi_* \leq \frac{Q(S,S^c)}{\pi(S)} = \frac{1}{N} \sum_{j=1}^{N} \frac{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}} \quad (35)$$

So we know $\gamma \leq 2\frac{1}{N} \sum_{j=1}^{N} \frac{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}}$. From a well-known theorem for the relationship between $t_{mix}(\epsilon)$ and γ: $t_{mix}(\epsilon) \geq (\frac{1}{\gamma} - 1) \log(\frac{1}{2\epsilon})$ [14], our conclusion follows. ∎

Theorem VII.4. *For an integer least-squares problem $\min_{\mathbf{s}\epsilon\{-1,+1\}^N} \|\mathbf{y} - \acute{\mathbf{H}}\mathbf{s}\|^2$, where $\acute{\mathbf{H}}$ is fixed and no two vectors give the same objective distance, the relaxation time (the inverse of the spectral gap) of the Markov chain for the reversible MCMC detector (Algorithm 5) is upper bounded by a constant as the temperature $\alpha \to 0$ if and only if there is no local minimum. Moreover, when there is a local minimum, as $\alpha \to 0$, the mixing time of Markov chain $t_{mix}(\epsilon) = e^{\Omega(\frac{1}{2\alpha^2})}$.*

Remarks: For the signal model $\mathbf{y} = \sqrt{\frac{\text{SNR}}{N}}\mathbf{H}\mathbf{x} + \boldsymbol{v}$, if α is set equal to the noise variance as in [15], [16], it is equivalent to setting "$\alpha \to 0$" when SNR $\to \infty$, since in Theorem VII.4 the SNR term is incorporated into $\acute{\mathbf{H}}$ and we keep $\acute{\mathbf{H}}$ fixed in Theorem VII.4.

Proof: First, when there is a local minimum, from Theorem VII.3 and Theorem VII.2, the spectral gap γ is lower bounded by

$$\gamma = \frac{2}{N} \sum_{j=1}^{N} \frac{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}} \quad (36)$$

As the temperature $\alpha \to 0$, the spectral gap upper bound

$$\frac{2}{1 + e^{\frac{\min_j \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2 - \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}}} \tag{37}$$

decreases at the speed of $\Theta(e^{-\frac{\min_j \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}_j\|^2 - \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2}{2\alpha^2}})$. So the relaxation time of the MCMC is lower bounded by $t_{mix}(\epsilon) = e^{\Omega(\frac{1}{2\alpha^2})}$, which grows unbounded as $\alpha \to 0$.

Suppose instead that there is no local minimum. We argue that as $\alpha \to 0$, the spectral gap of this MCMC is lower bounded by some constant independent of α. Again, we look at the bottle neck ratio and use Theorem VII.2 to bound the spectral gap.

Consider any set S of sequences which do not include the global minimum point $\mathbf{x}^*$. As $\alpha \to 0$, the measure of this set of sequences $\pi(S) \leq \frac{1}{2}$. Moreover, as $\alpha \to 0$, any set S with $\pi(S) \leq \frac{1}{2}$ can not contain the global minimum point $\mathbf{x}^*$. Now we look at the sequence $\tilde{\mathbf{x}}'$ which has the smallest distance $\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}'\|$ among the set S. Since there is no local minimum, $\tilde{\mathbf{x}}'$ must have at least one neighbor $\tilde{\mathbf{x}}''$ in S^c which has smaller distance than $\tilde{\mathbf{x}}'$. Otherwise, this would imply $\tilde{\mathbf{x}}'$ is a local minimum. So

$$Q(S, S^c) \geq \pi(\tilde{\mathbf{x}}') \times \frac{1}{N} \frac{e^{-\frac{\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}''\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}''\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2}{2\alpha^2}}} \tag{38}$$

As $\alpha \to 0$, $\frac{\pi(\tilde{\mathbf{x}}')}{\pi(S)} \to 1$. So for a given $\epsilon > 0$, as $\alpha \to 0$

$$\frac{Q(S, S^c)}{\pi(S)} \geq \frac{1 - \epsilon}{N} \frac{e^{-\frac{\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}''\|^2}{2\alpha^2}}}{e^{-\frac{\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}''\|^2}{2\alpha^2}} + e^{-\frac{\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2}{2\alpha^2}}}, \tag{39}$$

which approaches $\frac{(1-\epsilon)}{N}$ as $\alpha \to 0$ because $\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}''\|^2 < \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2$.

From Theorem VII.2, the spectral gap γ is at least $\frac{(\frac{Q(S,S^c)}{\pi(S)})^2}{2}$, which is lower bounded by a constant as $\alpha \to 0$. ∎

So from the analysis above, the mixing time is closely related to whether there are local minima in the problem. In the next section, we will see there often exist local minima, which implies very slow convergence rate for MCMC when the temperature is kept at the noise level in the high SNR regime.

VIII. PRESENCE OF LOCAL MINIMA

We have seen that the mixing time of MCMC detectors are closely related to the existence of local minima. It is natural to ask how often local minima occur in ILS problems. In this section, we derive some results about how many local minima there are in an ILS problem, especially when the SNR is high.

Theorem VIII.1. *There can be exponentially many local minima in an integer least-quare problem*

Proof: See Appendix XI-B for a detailed proof. ∎

Now we study how often we encounter a local minimum in specific ILS problem models. Without loss of generality, we assume that the transmitted sequence is an all $-\mathbf{1}$ sequence.

We first give the condition for $\tilde{\mathbf{x}}$ to be a local minimum. We assume that $\tilde{\mathbf{x}}$ is a vector which has k '+1' over an index set K with $|K| = k$ and $(N-k)$ '−1' over the set $\overline{K} = \{1, 2, ..., N\} \smallsetminus K$.

Lemma VIII.2. *$\tilde{\mathbf{x}}$ is a local minimum if and only if $\tilde{\mathbf{x}}$ is not a global minimum; and*

- $\forall i \in K$,

$$\mathbf{h}_i^T \left(\sum_{j \in K} \mathbf{h}_j - \frac{\boldsymbol{v}}{2} \right) < \frac{\|\mathbf{h}_i\|^2}{2} \tag{40}$$

- $\forall i \in \overline{K}$,

$$\mathbf{h}_i^T \left(\sum_{j \in K} \mathbf{h}_j - \frac{\boldsymbol{v}}{2} \right) > -\frac{\|\mathbf{h}_i\|^2}{2}. \tag{41}$$

Proof: For a position $i \in K$, when we flip $\tilde{\mathbf{x}}_i$ to 1, $\|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2$ is increased, namely,

$$\begin{aligned} & \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2 - \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}_{\sim i}\|^2 \\ = & \| -2 \sum_{j \in K} \mathbf{h}_j + \boldsymbol{v}\|^2 - \| -2 \sum_{j \in K, j \neq i} \mathbf{h}_j + \boldsymbol{v}\|^2 \\ = & 4\|\mathbf{h}_i\|^2 + 4\mathbf{h}_i^T \left(2 \sum_{j \in K, j \neq i} \mathbf{h}_j - \boldsymbol{v} \right) \\ < & 0, \end{aligned} \tag{42}$$

where $\tilde{\mathbf{x}}_{\sim i}$ is a neighbor of $\tilde{\mathbf{x}}$ by changing index i. This means

$$\mathbf{h}_i^T \left(\sum_{j \in K} \mathbf{h}_j - \frac{\boldsymbol{v}}{2} \right) < \frac{\|\mathbf{h}_i\|^2}{2}. \tag{43}$$

For a position $i \in \overline{K}$, when we flip $\tilde{\mathbf{x}}_i$ to −1, $\|\mathbf{y} - \mathbf{h}\tilde{\mathbf{x}}'\|^2$ is also increased, namely,

$$\begin{aligned} & \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}\|^2 - \|\mathbf{y} - \acute{\mathbf{H}}\tilde{\mathbf{x}}_{\sim i}\|^2 \\ = & \| -2 \sum_{j \in K} \mathbf{h}_j + \boldsymbol{v}\|^2 - \| -2 \sum_{j \in K} \mathbf{h}_j - 2\mathbf{h}_i + \boldsymbol{v}\|^2 \\ = & -4\|\mathbf{h}_i\|^2 + 4\mathbf{h}_i^T \left(-2 \sum_{j \in K} \mathbf{h}_j + \boldsymbol{v} \right) \\ < & 0. \end{aligned} \tag{44}$$

This means

$$(\mathbf{h}_i)^T \left(\sum_{j \in K} \mathbf{h}_j - \frac{\boldsymbol{v}}{2} \right) > -\frac{\|\mathbf{h}_i\|^2}{2}. \tag{45}$$

∎

It is not hard to see that when SNR $\to \infty$, $\boldsymbol{v}$ is comparatively small with high probability, so we have the following lemma.

Lemma VIII.3. *When $SNR \to \infty$, $\tilde{\mathbf{x}}$ is a local minimum with high probability, if and only if $\tilde{\mathbf{x}} \neq -\mathbf{1}$; and*

- $\forall i \in K$,

$$\mathbf{h}_i^T \left(\sum_{j \in K} \mathbf{h}_j \right) < \frac{\|\mathbf{h}_i\|^2}{2} \tag{46}$$

- $\forall i \in \overline{K}$,

$$\mathbf{h}_i^T \left(\sum_{j \in K} \mathbf{h}_j \right) > -\frac{\|\mathbf{h}_i\|^2}{2}. \tag{47}$$

We now set out to investigate the chance of having a local minimum in MIMO systems.

Theorem VIII.4. *Consider a 2×2 matrix $\acute{\mathbf{H}}$ whose two columns are uniform randomly sampled from the unit-normed 2-dimensional vector. When $\boldsymbol{v} = 0$, the probability of there existing a local minimum for such an $\acute{\mathbf{H}}$ is $\frac{1}{3}$.*

Please see the appendix for its proof.

Theorem VIII.5. *Consider a 2×2 matrix $\acute{\mathbf{H}}$ whose elements are independent $\mathcal{N}(0,1)$ Gaussian random variables. When $\boldsymbol{v} = 0$, the probability of there existing a local minimum for such an $\acute{\mathbf{H}}$ is $\frac{1}{3} - \frac{1}{\sqrt{5}} + \frac{2\arctan(\sqrt{\frac{5}{3}})}{\sqrt{5}\pi}$.*

Please refer to the appendix for its proof.

For higher dimension N, it is hard to directly estimate the probability of a vector being a local minimum based on the conditions in Lemma VIII.2. Simulation results instead suggest that for large N, with high probability, there exists at least one local minimum. We conjecture this is the case, but proof or disproof of it seems nontrivial.

IX. Choice of Temperature α in High SNR

In previous sections, we have looked at the mixing time of MCMC for an ILS problem. Now we use the results we have accumulated so far to help choose the appropriate temperature of α to ensure that the MCMC mixes fast and that the optimal solution also comes up fast when the system is in a stationary distribution.

When SNR $\to \infty$, the ILS problem will have the same local minima as the case $\boldsymbol{v} = 0$. From the derivations and simulations, it is suggested that with high probability there will be at least one local minimum, especially for large problem dimension N.

So following from Lemma VII.4 and the reasoning therein, to ensure there is an upper bound on the mixing time as SNR $\to \infty$, the temperature α should at least grow at a rate such that

$$\max_{\tilde{\mathbf{x}}} \min_{\tilde{\mathbf{x}}'} \frac{\frac{\text{SNR}}{N} \left(\| - \acute{\mathbf{H}}\mathbf{1} - \tilde{\mathbf{x}}' \|^2 - \| - \mathbf{H}\mathbf{1} - \tilde{\mathbf{x}} \|^2 \right)}{2\alpha^2} \le C, \quad (48)$$

where $\tilde{\mathbf{x}}$ is a local minimum and $\tilde{\mathbf{x}}'$ is a neighbor of $\tilde{\mathbf{x}}$, and C is a constant.

This requires that α^2 grow as fast as $\Omega(\text{SNR})$ to ensure fast mixing with the existence of local minima. This explains that if we keep the temperature at the noise level, it will lead to slow convergence in the high SNR regime [17].

We remark that, for the squared-norm-2 MCMC detector, it is hard to explicitly evaluate α from (48). However, for the norm-2 MCMC detector, the corresponding criterion for a fast mixing is given by

$$\max_{\tilde{\mathbf{x}}} \min_{\tilde{\mathbf{x}}'} \frac{\sqrt{\frac{\text{SNR}}{N}} \left(\| - \mathbf{H}\mathbf{1} - \tilde{\mathbf{x}}' \|_2 - \| - \mathbf{H}\mathbf{1} - \tilde{\mathbf{x}} \|_2 \right)}{2\alpha^2} \le C, \quad (49)$$

for some constant C.

Table I: Theoretical values of α for $N = 10$.

SNR	10 dB	14 dB
$\alpha_{+,\zeta=2}$	4.98	7.99
$\alpha_{+,\zeta=1}$	3.54	5.76
$\tilde{\alpha}_+$	2.74	4.56

By the triangular inequality, and the concentration of measure result for Gaussian random variables, with high probability as $N \to \infty$, for any $\epsilon > 0$,

$$\max_{\tilde{\mathbf{x}}} \min_{\tilde{\mathbf{x}}'} \sqrt{\frac{\text{SNR}}{N}} \left(\| - \mathbf{H}\mathbf{1} - \tilde{\mathbf{x}}' \|_2 - \| - \mathbf{H}\mathbf{1} - \tilde{\mathbf{x}} \|_2 \right)$$
$$\le (1+\epsilon)2\sqrt{\text{SNR}}. \quad (50)$$

So for the norm-2 MCMC detector, as long as $\alpha^2 \ge (1+\epsilon)\sqrt{\text{SNR}}/C$, the condition (49) will be satisfied with high probability.

X. Simulation Results

In this section we present simulation results for an $N \times N$ MIMO system with a full square channel matrix containing i.i.d. Gaussian entries. In Figure 2 and Figure 3 the Bit Error Rate (BER) of the sequential MCMC detector has been evaluated as a function of the number of block iterations in a 10×10 system using a varity of α values. Thereby, we can inspect how the parameter α affects the convergence rate of the MCMC detector and, as a reference, we have included the values of α computed using (23) and (26), which can be seen in Table I.

The performance of the Maximum Likelihood (ML), the Zero-Forcing (ZF), and the Linear Minimum Mean Square Error (LMMSE) detector have also been plotted, to ease the comparison of the MCMC detector with these detectors (Please see [3], for example, for descriptions of these well-known detectors). It is seen that the MCMC detector outperforms both the ZF and the LMMSE detector after only a few block iterations in all the presented simulations, when the tuning parameter α is chosen properly. Furthermore, it is observed that the parameter α has a huge influence on the convergence rate and that the MCMC detector converges toward the ML solution as a function of the iterations[3]. Figure 4 shows the BER performance for the MCMC detector for fixed number of iterations, $k = 100$. From the figure we see that the SNR has a significant influence on the optimal choice of α given a fixed number of iterations. The performance of the sequential MCMC detector is also shown for a 50×50 system, which represents a ML decoding problem of huge complexity where an exhaustive search would require $2^{50} \approx 10^{15}$ evaluations. For this problem even the sphere decoder would have an enormous complexity under moderate SNR, and it has actually been proved in [9] that the complexity of SD for SNR $= O(\ln(N))$ is exponential. Therefore, it has not been possible to simulate the performance of this decoder within a reasonable time

---

[3]It should be noted that the way we decode the symbol vector to a given iteration, is to select the symbol vector with has the lowest cost function in all the iterations up to that point in time.
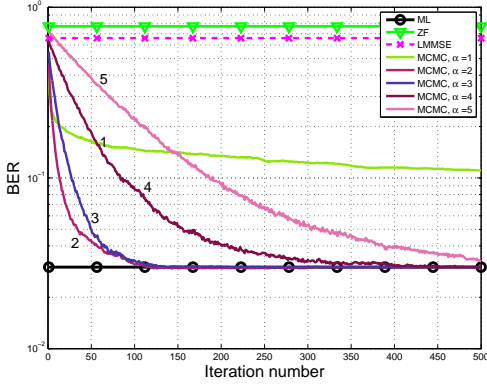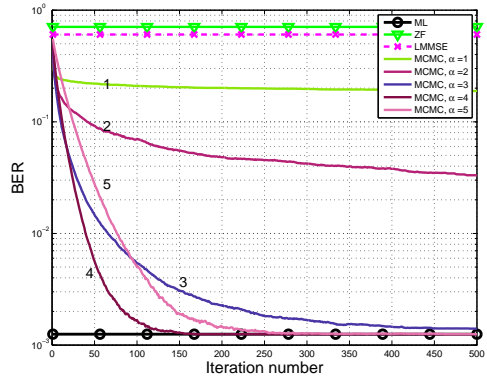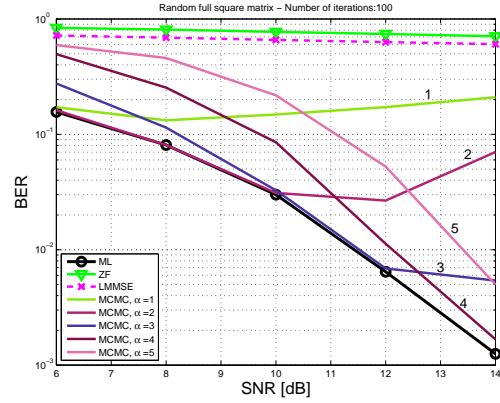
Figure 2: BER vs. iterations, 10×10. SNR = 10 dB.

Figure 4: BER vs. SNR, 10×10. Number of iterations, $k = 100$.

Figure 3: BER vs. iterations, 10×10 system. SNR = 14 dB.
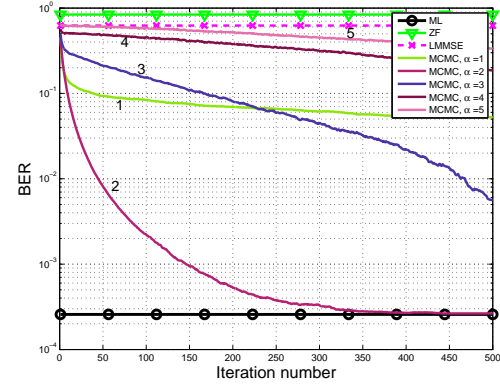

Figure 5: BER vs. iterations, 50×50 system. SNR = 12 dB.

and we have therefore "cheated" a little by initializing the radius of the sphere to the minimum of either the norm of the transmitted symbol vector or the solution found by the MCMC detector. This has been done in order to evaluate the BER performance of the optimal detector. Figure 5 shows the BER curve as a function of the iteration number while Figure 6 illustrates the BER curve vs. the SNR. From Figure 5 we see that there is a quite good correspondence between the simulated α and the theoretical value $\tilde{\alpha}_+ = 2.6$ obtained from (26).

Now we compare the numerical performance of reversible MCMC detectors including square-norm-2 MCMC detector and norm-2 MCMC detector. Again, we simulate $N \times N$ MIMO systems with channel matrices containing zero mean i.i.d Gaussian entries. Figure 7 shows the BER as a function of SNR for two different MCMC detectors: squared-norm-2 and norm-2 MCMC detector, when $N = 10$. 1000 iterations are used in both reversible MCMC detectors which are initialized with a random input vector. Dashed and blue curves in Figure 7 represent squared-norm-2 MCMC detector for various α values. Squared-norm-2 MCMC detector uses $\|y - \sqrt{\frac{\text{SNR}}{\text{N}}}\mathbf{H}\hat{s}\|^2$ in the calculation of the probability of transferring from one state to another. Solid and red curves represent norm-2 MCMC detector which uses equation $\|y - \sqrt{\frac{\text{SNR}}{\text{N}}}\mathbf{H}\hat{s}\|$. We can see that

for the same α values, norm-2 MCMC detector has better BER compared with the Squared-norm-2 in high SNR.

Now we consider numerical results related to the mixing time of reversible MCMC detectors. In Figure 8, we plot the expected number of local minima in a system as the problem dimension N grows. For each N, we generate 100 random channel matrices and for each matrix, we examine the number of local minima by exhaustive search. As the problem dimension N grows, the number of local minima
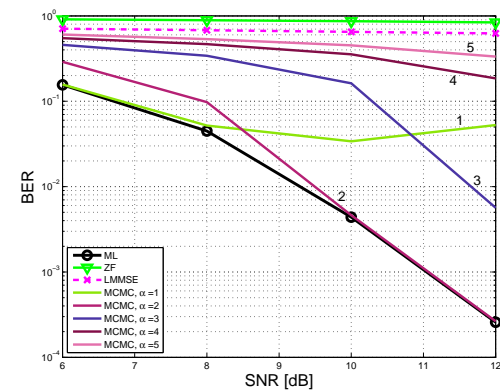

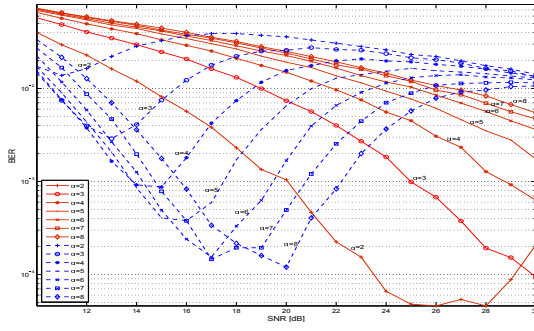Figure 6: BER vs. SNR, 50×50 system. Number of iterations, $k = 500$.

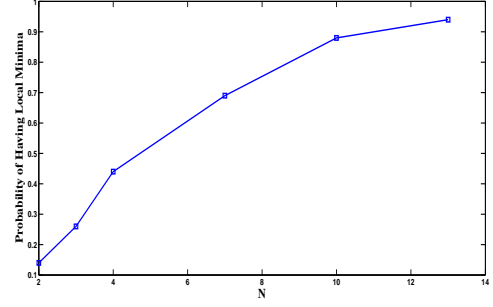Figure 7: BER vs SNR for squared-norm-2 MCMC detector (dashed lines) and norm-2 MCMC detector (solid lines)
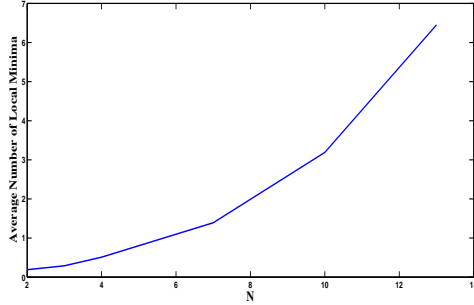

Figure 9: The probability of having local minima

increases with the increasing of the number of local minima.

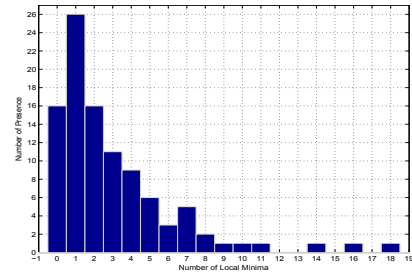Figure 8: Average number of local minima

Figure 10: Histograms of the number of local minima for N=10

grows rapidly.

In Figure 9, we plot the probability of there existing a local minimum as the problem dimension N grows. For each N, we generated 100 random channel matrices and for each matrix, we examined whether there exist local minima by exhaustive search. As N grows, the empirical probability of there existing at least one local minimum approaches 1. It is interesting to see that for $N = 2$, our theoretical result $\frac{1}{3} - \frac{1}{\sqrt{5}} + \frac{2\arctan(\sqrt{\frac{5}{3}})}{\sqrt{5}\pi} \approx$ 0.15 matches well with the simulations.

Figures 10 and 11 show the histograms of the number of local minima for $N = 10$ and 12 respectively, under SNR = 10. For each parameter N, we used exhaustive search to examine the number of local minima in 100 randomly chosen Gaussian channel matrices. Obviously, the average number of local minima increases as N increases, while the frequency of 0 local minima decreases.

Figure 12 presents the histograms of the spectral gap when there are 0, 1, 2, and 3 local minima respectively for $N = 5$ and SNR = 10. We generated 10^5 randomly Gaussian channel matrices. In each matrix we examined the number of local minima and calculated the spectral gap when $\alpha^2 = 1$. For all these figures, each bar represents the percentage of matrices which fall in a spectral gap interval of 0.01. We can see that, when there is 0 local minimum, around 50 percent of the matrices' spectral gap fall between 0.19 and 0.2, suggesting these MCMC detectors mix fast. However, when there is at least one local minimum, a high percentage of the matrices have spectral gap values between 0 and 0.01. This percentage
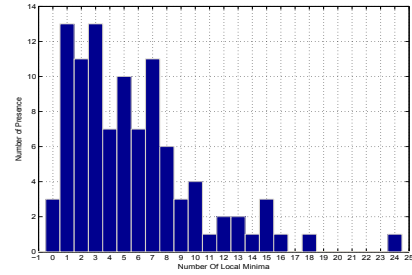

Figure 11: Histograms of the number of local minima for N=12

XI. APPENDIX

A. Proving Lemma IV.1

Lemma IV.1 (Gaussian Integral) *Let* $\mathbf{v}$ *and* $\mathbf{x}$ *be independent Gaussian random vectors with distribution* $\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ *each. Then*

$$\mathcal{E}\left\{e^{\eta(\|\mathbf{v}+a\mathbf{x}\|^2 - \|\mathbf{v}\|^2)}\right\} = \left(\frac{1}{1 - 2a^2\eta(1+2\eta)}\right)^{N/2}. \quad (51)$$

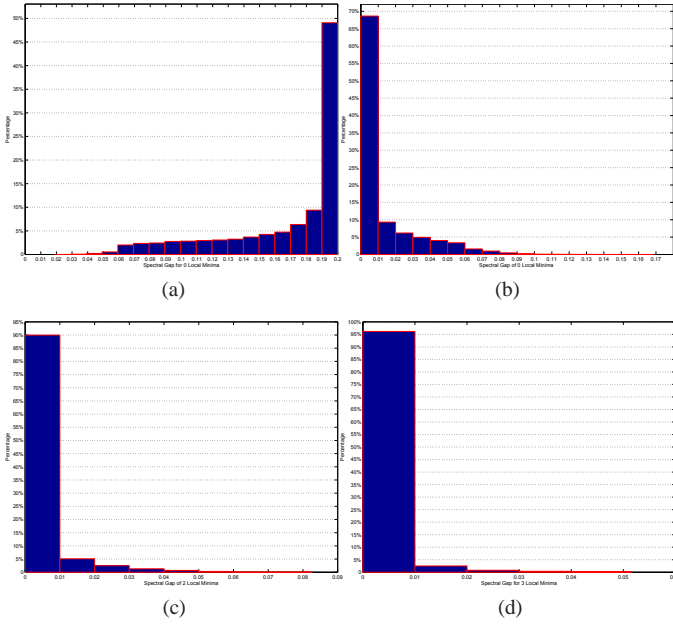**Proof:** In order to determine the expected value we compute

Figure 12: Spectral gap with (a) 0 (b) 1 (c) 2 (d) 3 local minima

the multivariate integral

$$\mathcal{E}\left\{e^{\eta\left(\|\mathbf{v}+a\mathbf{x}\|^2-\|\mathbf{v}\|^2\right)}\right\} \tag{52a}$$

$$= \int \frac{d\mathbf{x}d\mathbf{v}}{(2\pi)^N} e^{-\frac{1}{2}\left[\mathbf{v}^T,\ \mathbf{x}^T\right]\begin{bmatrix} \mathbf{I}_N & -2a\eta\mathbf{I}_N \\ -2a\eta\mathbf{I}_N & (1-2a^2\eta)\mathbf{I}_N \end{bmatrix}\begin{bmatrix} \mathbf{v} \\ \mathbf{x} \end{bmatrix}} \tag{52b}$$

$$= \frac{1}{\det^{1/2}\begin{bmatrix} \mathbf{I}_N & -2a\eta\mathbf{I}_N \\ -2a\eta\mathbf{I}_N & (1-2a^2\eta)\mathbf{I}_N \end{bmatrix}} \tag{52c}$$

$$= \frac{1}{\det^{N/2}\begin{bmatrix} 1 & -2a\eta \\ -2a\eta & 1-2a^2\eta \end{bmatrix}} \tag{52d}$$

$$= \left(\frac{1}{1-2a^2\eta(1+2\eta)}\right)^{N/2}. \tag{52e}$$

Thus, Lemma IV.1 has hereby been proved. ∎

B. Proving Lemma VIII.1

Proof: Let N be an even integer. Consider a matrix whose first $\frac{N}{2}$ columns $\mathbf{h}_i$, $1 \le i \le \frac{N}{2}$ have unit norms and are orthogonal to each other. For the other $\frac{N}{2}$ columns $\mathbf{h}_i$, $\frac{N}{2}+1 \le i \le N$, $\mathbf{h}_i = -(1+\epsilon)\mathbf{h}_{i-\frac{n}{2}}$, where ϵ is a sufficiently small positive number ($\epsilon < 1$). We also let $\mathbf{y} = \acute{\mathbf{H}}(-\mathbf{1})$, where $\mathbf{1}$ is an all-1 vector. So $-\mathbf{1}$ is a globally minimum point for this ILS problem.

Consider all those vectors $\tilde{\mathbf{x}}'$ which, for any $1 \le i \le \frac{N}{2}$, its i-th element and $i+\frac{N}{2}$-th element are either simultaneously +1 or simultaneously −1. When ϵ is smaller than 1, we claim that any such a vector except the all −1 vector $\tilde{\mathbf{x}}$, is a local

minimum, which shows that there are at least $2^{\frac{N}{2}} - 1$ local minima.

Assume that for a certain $1 \le i \le \frac{N}{2}$, the i-th element and $(i+\frac{N}{2})$-th element of $\tilde{\mathbf{x}}'$ are simultaneously −1. Then if we change the i-th element to +1, $\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2$ increases by 4; and if we change the $(i+\frac{N}{2})$-th element to +1, $\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2$ increases by $4(1+\epsilon)^2$. This is true because the i-th and $(i+\frac{N}{2})$-th columns are orthogonal to other $(N-2)$ columns.

Similarly, assume that for a certain $1 \le i \le \frac{N}{2}$, the i-th element and $(i+\frac{N}{2})$-th element of $\tilde{\mathbf{x}}'$ are simultaneously +1. Then if we change the i-th element to −1, $\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2$ increases by $4(1+\epsilon)^2 - 4\epsilon^2$; and if we change the $(i+\frac{N}{2})$-th element to −1, $\|\mathbf{y}-\acute{\mathbf{H}}\tilde{\mathbf{x}}'\|^2$ increases by $4 - 4\epsilon^2$. ∎

C. Proving Lemma VIII.4

Proof: When $\boldsymbol{v} = 0$, clearly $\tilde{\mathbf{x}} = (-1,-1)$ is a global minimum point, not a local minimum point. It is also clear that $\tilde{\mathbf{x}} = (-1,1)$ or $\tilde{\mathbf{x}} = (1,-1)$ can not be a local minimum point since they are neighbors to the global minimum solution. So the only possible local minimum point is $\tilde{\mathbf{x}} = (1,1)$.

From Lemma VIII.2, the corresponding necessary and sufficient condition is

$$\mathbf{h}_1^T\mathbf{h}_2 < -\frac{\|\mathbf{h}_1\|^2}{2} = -\frac{\|\mathbf{h}_2\|^2}{2} = -\frac{1}{2}.$$

This means the angle θ between the two 2-dimensional vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ satisfy $\cos(\theta) < -\frac{1}{2}$. Since $\mathbf{h}_1$ and $\mathbf{h}_2$ are two independent uniform randomly sampled vector, the chance for that to happen is $\frac{\pi-\arccos\left(-\frac{1}{2}\right)}{\pi} = \frac{1}{3}$. ∎

D. Proving Lemma VIII.5

Proof: When $\boldsymbol{v} = 0$, clearly $\tilde{\mathbf{x}} = (-1,-1)$ is a global minimum point, not a local minimum point. It is also clear that $\tilde{\mathbf{x}} = (-1,1)$ or $\tilde{\mathbf{x}} = (1,-1)$ can not be a local minimum point since they are neighbors to the global minimum solution. So the only possible local minimum point is $\tilde{\mathbf{x}} = (1,1)$.

From Lemma VIII.2, the corresponding necessary and sufficient condition is

$$\mathbf{h}_1^T\mathbf{h}_2 < -\max\left\{\frac{\|\mathbf{h}_1\|^2}{2}, \frac{\|\mathbf{h}_2\|^2}{2}\right\}.$$

This means the angle θ between the two 2-dimensional vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ satisfy

$$r_1r_2\cos(\theta) < -\frac{\max\left\{r_1^2, r_2^2\right\}}{2},$$

where r_1 and r_2 are respectively the ℓ_2 norm of $\mathbf{h}_1$ and $\mathbf{h}_2$.

Because the elements of $\acute{\mathbf{H}}$ are independent Gaussian random variables, r_1 and r_2 are thus independent random variables following the Rayleigh distribution

$$p(r_1) = r_1e^{-\frac{r_1^2}{2}}, p(r_2) = r_2e^{-\frac{r_2^2}{2}},$$

while θ follows a uniform distribution over $[0, 2\pi)$

By symmetry, for $t \geq 1$,

$$P(\frac{\max\{r_1^2, r_2^2\}}{r_1 r_2} > t)$$

$$= 2 \int_0^\infty r_1 e^{-\frac{r_1^2}{2}} \times \int_0^{\frac{r_1}{t}} r_2 e^{-\frac{r_2^2}{2}} \, dr_2 \, dr_1$$

$$= 2 \int_0^\infty r_1 e^{-\frac{r_1^2}{2}} \times (1 - e^{-\frac{r_1^2}{2}}) \, dr_1$$

$$= 2(1 - \int_0^\infty r_1 e^{-(\frac{1}{2} + \frac{1}{2t^2})r_1^2} \, dr_1)$$

$$= \frac{2}{t^2 + 1}.$$

Since θ is an independent random variable satisfying $\cos(\theta) < -\frac{\max\{r_1^2, r_2^2\}}{2r_1 r_2}$ and $\cos(\theta) \geq -1$, the probability that $\tilde{\mathbf{x}} = (+1, +1)$ is a local minimum is given by

$$P = \int_1^2 (1 - \frac{2}{t^2 + 1})'(1 - \frac{\arccos(-\frac{t}{2})}{\pi}) \, dt$$

$$= \int_1^2 \frac{4t}{(t^2 + 1)^2}(1 - \frac{\arccos(-\frac{t}{2})}{\pi}) \, dt.$$

$$= \frac{1}{3} - \frac{1}{\sqrt{5}} + \frac{2 \arctan(\sqrt{\frac{5}{3}})}{\sqrt{5}\pi},$$

which is approximately 0.145696. ∎

REFERENCES

[1] M. Hansen, B. Hassibi, A. Dimakis, and W. Xu, "Near-Optimal Detection in MIMO Systems using Gibbs Sampling ," in *Globecom'09*, 2009.
[2] W. Xu, A. Dimakis, and B. Hassibi, "On the mixing time of markov chain monte carlo for integer least-square problems," in *IEEE 51st Annual Conference on Decision and Control (CDC)*, 2012, pp. 2545–2550.
[3] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.
[4] M. A. Borno, "Reduction in Solving Some Integer Least Squares Problems," *Thesis, McGill University*, 2011.
[5] M. O. Damen, H. E. Gamal, and G. Caire, "On Maximum-Likelihood Detection and the Search for the Closest Lattice Point," *IEEE Trans. on Info. Theory*, vol. 49, pp. 2389–2402, Oct. 2003.
[6] B. M. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Commun.*, vol. 51, no. 3, pp. 389–399, 2003.
[7] B. Hassibi and H. Vikalo, "On the Sphere-Decoding Algorithm. I. Expected Complexity," *IEEE Trans. on Sig. Proc.*, vol. 53, pp. 2806–2818, Aug. 2005.
[8] ——, "On the Sphere-Decoding Algorithm. II. Generalizations, Second-Order Statistics, and Applications to Communications," *IEEE Trans. on Sig. Proc.*, vol. 53, pp. 2819–2834, Aug. 2005.
[9] J. Jaldén and B. Ottersten, "On the Complexity of Sphere Decoding in Digital Communications," *IEEE Trans. on Sig. Proc.*, vol. 53, pp. 1474–1484, Apr. 2005.
[10] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
[11] T. Datta, N. Srinidhi, A. Chockalingam, and B. Rajan, "Random-restart reactive tabu search algorithm for detection in large-MIMO systems," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1107–1109, 2010.
[12] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer, 2004.
[13] O. Häggström, *Finite Markov chains and algorithmic applications*. Cambridge University Press, 2002.
[14] D. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
[15] X. Wang and V. H. Poor, *Wireless Communications Systems: Advanced Techniques for Signal Reception*. Prentice Hall, 2003.
[16] H. Zhu, B. Farhang-Boroujeny, and R. Chen, "On performance of sphere decoding and Markov chain Monte Carlo detection methods," *IEEE Sig. Proc. Letters*, vol. 12, pp. 669–672, 2005.
[17] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Trans. on Sig. Proc.*, vol. 54, no. 5, pp. 1896–1909, 2006.
[18] R. Chen, S. J. Liu, and X. Wang, "Convergence analyses and comparisons of Markov chain Monte Carlo algorithms in digital communications," *IEEE Transactions on Signal Processing*, vol. 50, pp. 255–270, 2002.
[19] D. MacKay, *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.
[20] M. Jerrum and A. Sinclair, "Approximating the Permanent," *SIAM Journal on Computing*, vol. 18, pp. 1149–1178, 1989.
[21] G. Lawler and A. Sokal, "Bounds on the L^2 spectrum for Markov Chains and Markov Processes: a Generalization of Cheeger's Inequality," *Trans. Amer. Math. Soc.*, vol. 309, pp. 557–580, 1988.