A new theorem shows that noise can speed convergence to equilibrium in discrete finite-state Markov chains. The noise applies to the state density and helps the Markov chain explore improbable regions of the state space. The theorem ensures that a stochastic-resonance noise benefit exists for states that obey a vector-norm inequality. Such noise leads to faster convergence because the noise reduces the norm components. A corollary shows that a noise benefit still occurs if the system states obey an alternate norm inequality. This leads to a noise-benefit algorithm that requires knowledge of the steady state. An alternative blind algorithm uses only past state information to achieve a weaker noise benefit. Simulations illustrate the predicted noise benefits in three well-known Markov models. The first model is a two-parameter Ehrenfest diffusion model that shows how noise benefits can occur in the class of birth-death processes. The second model is a Wright-Fisher model of genotype drift in population genetics. The third model is a chemical reaction network of zeolite crystallization. A fourth simulation shows a convergence rate increase of 64% for states that satisfy the theorem and an increase of 53% for states that satisfy the corollary. A final simulation shows that even suboptimal noise can speed convergence if the noise applies over successive time cycles. Noise benefits tend to be sharpest in Markov models that do not converge quickly and that do not have strong absorbing states.

I. NOISE BENEFITS IN MARKOV CHAIN DENSITY ESTIMATION

We show that noise can speed convergence to equilibrium in discrete finite-state Markov chains. Judiciously adding noise directly to the state density speeds up the convergence time for the Markov chain simulation depending on the direction of an inequality. The noise appears to give the Markov chain system access to a statistically richer set of otherwise improbable states. Neural network researchers have observed related smoothing effects that arise when adding noise to training data [1, 2]. Figures 1, 2, and 7 show a Markov chain noise benefit when a simulation adds noise to the current state density. Table II summarizes the 53% – 64% performance improvement after adding noise in a simulation. Figure 10 shows that this noise benefit holds over successive cycles.

Many nonlinear signal systems benefit from adding small amounts of noise [3–16]. Too little noise produces little or no benefit while too much noise can swamp the system’s performance. This so-called “stochastic resonance” effect can take the form of an increased signal-to-noise ratio [17–19], entropy-based bit count [20–22], input-output correlation [23], or probability of detection [24, 25]. The noise benefit for a simulated Markov chain is a shorter time to converge to the equilibrium probability density in the sense that the noise reduces the vector norm of the error.

Markov chains form a basis for powerful Markov chain Monte Carlo (MCMC) statistical simulations [26]. MCMC methods generate samples from a given posterior probability density function by constructing a Markov chain whose stationary density equals the posterior of interest [27, 28]. The Metropolis-Hastings algorithm [29, 30] and Gibbs samplers [31, 32] are special and powerful MCMC frameworks that compute Bayesian statistics. But MCMC methods suffer from problem-specific parameters that govern sample acceptance and convergence assessment [33, 34]. Strong dependence on initial conditions also biases the MCMC sampling unless the simulation allows a lengthy period of “burn-in” to allow the driving Markov chain to mix adequately [26, 35].

The Markov Chain Noise Benefit Theorem in the next section shows how to construct a normalized state density at each time cycle for a finite time-homogeneous Markov chain with an irreducible and aperiodic state transition matrix. The theorem and corollary guarantee the existence of a component-wise noise benefit that decreases the time to convergence. They show that simulations can perturb the current state of a Markov chain to explore novel regions in the state space and speed convergence to the steady-state distribution. The form of the noise depends on the direction of a state-related inequality. The theorem may ensure only minimal benefits for systems that exhibit fast convergence or that possess strong absorbing states.

Section III presents two algorithms that use the Markov Chain Noise Benefit Theorem to obtain a noise benefit. The first algorithm shows how the simulation can obtain an optimal noise benefit. The second algorithm describes how to obtain a noise benefit that uses
II. MARKOV CHAIN NOISE BENEFIT THEOREM

The Markov Chain Noise Benefit Theorem below shows that Markov chain simulations can benefit from noise through faster convergence. Markov chain simulations employ a stochastic discrete time model to estimate the probability density over a system’s state-space.

Suppose $M$ is a time-homogeneous Markov chain over a finite state-space with $N < \infty$ states [26, 36, 37]. Let the $N \times 1$ column-vector $x(t)$ represent the state of the Markov chain at time $t$. Each component $x_i(t)$ represents the probability that the chain is in the corresponding state $i$ at time $t$. Then

$$\sum_{i=1}^{N} x_i(t) = 1$$

(1)

for all $t$ because $x(t)$ is a probability density over the $N$ states.

Let the $P$ represent the single-step state transition probability matrix where

$$P_{i,j} = P[x_{t+1} = j \mid x_t = i]$$

(2)

is the probability of the chain in state $i$ at time $t$ moving to state $j$ at time $t + 1$. Then there exists a stationary vector $x^\infty$ such that

$$x^\infty = x^\infty P$$

(3)

[37]. So $x^\infty$ is always a left eigenvector of the transition probability matrix $P$ that corresponds to the eigenvalue $\lambda = 1$.

The $n$-step transition probability matrix $P^{(n)}$ has entries

$$P^{(n)}_{i,j} = P[X_{t+n} = j \mid X_t = i]$$

(4)

$$= \sum_{k=0}^{N} P[X_{t+n} = j \mid X_t = i, X_{t+1} = k] \cdot P[X_{t+1} = k \mid X_t = i]$$

(5)

$$= \sum_{k=0}^{N} P[X_{t+n} = j \mid X_{t+1} = k] \cdot P[X_{t+1} = k \mid X_t = i]$$

(6)

$$= \sum_{k=0}^{N} P^{(n-1)}_{k,j} P_{t,k}$$

(7)

where $P^{(n)}_{i,j}$ is the probability that the chain transitions from state $i$ to state $j$ in exactly $n$ time steps. State $j$ is accessible from state $i$ if there is some non-zero probability of transitioning from state $i$ to state $j$ in any number of steps:

$$P^{(n)}_{i,j} > 0$$

(8)

for some $n > 0$.

A Markov chain is irreducible if every state is accessible from every other state [26, 37]. Irreducibility implies that for all states $i$ and $j$ there exists $m > 0$ such that

$$P[X_{n+m} = j \mid X_n = i] = P^{(n)}_{i,j} > 0.$$  

(9)

This is equivalent to $P$ is a regular stochastic matrix if $M$ is a finite Markov chain.

The period $d_i$ of state $i$ is

$$d_i = \gcd \{ n \geq 1 : P^{(n)}_{i,i} > 0 \}$$

(10)

or $d_i = \infty$ if $P^{(n)}_{i,i} = 0$ for all $n \geq 1$ where $\gcd$ denotes the greatest common divisor. State $i$ is aperiodic if $d_i = 1$. A Markov chain with transition matrix $P$ is aperiodic if $d_i = 1$ for all states $i$.

Suppose a Markov chain $M$ is irreducible and aperiodic. Then the fixed point $x^\infty$ is unique and

$$\lim_{k \to \infty} P^{(k)} = 1 \otimes x^\infty$$

(11)

where $1$ is the column vector with all entries equal to 1 [36, 38]. The outer product generates a rank-one $N \times N$ matrix with each column equal to the stationary state density.

The Markov Chain Noise Benefit Theorem below shows that there is a component-wise noise benefit for any component that has not yet converged to its stationary value. The theorem assumes that the sign of a state-related inequality is in one of two directions. The corollary assumes it is in the other direction. The proof of the theorem is lengthy and appears in the Appendix.

Markov Chain Noise Benefit Theorem. Suppose $M$ is a finite time-homogeneous Markov chain with $N$ states and transition matrix $P$. Suppose further that $M$ is irreducible and aperiodic. Then for all non-stationary state density vectors $x$ there exists a noise benefit in the sense that there exists some $A > 0$ so that for all $a \in (0, A)$:

$$||\bar{x} P - x^\infty|| < ||x P - x^\infty||$$

(12)

for all states $i$ with

$$\Delta_i = (x - x^\infty) P_i > 0$$

(13)

where

$$\bar{x} = \frac{1}{1 + a} (x + n)$$

(14)
is the normalized state vector after adding a noise vector \( n \) with only one non-zero component:
\[
n_j = \begin{cases} 
  a & j = k \\
  0 & j \neq k
\end{cases}
\]  
(15)

for any \( k \) that satisfies
\[
\Delta_k = (x - x^\infty) P_k > 0.
\]  
(16)

The following corollary provides a complementary result when the converse of inequality (13) holds (\( \Delta_i < 0 \)) in the Markov Chain Noise Benefit Theorem.

**Corollary.** Suppose the hypotheses of the Markov Chain Noise Benefit Theorem hold. Then there exists a noise benefit for each non-stationary state density vector \( x \) in the sense that there exists some \( A > 0 \) so that for all \( a \in (0, A) \):
\[
||\tilde{x} P - x^\infty||_i < ||x P - x^\infty||_i
\]  
(17)

for all states \( i \) with
\[
\Delta_i = (x - x^\infty) P_i < 0.
\]  
(18)

**Proof.** The \( \Delta_i \) sign change does not affect the expansion.
in the proof of the theorem. So

\[ |\tilde{x}P - x^\infty| = |\Delta_i - \alpha_i \frac{1}{1 + \alpha_i} (xP_i - P_{k, i})| \quad (19) \]

holds. Now \( \Delta_i < 0 \) by hypothesis. Thus

\[ |\Delta_i - \alpha_i \frac{1}{1 + \alpha_i} (xP_i - P_{k, i})| < |\Delta_i| \quad (20) \]

if and only if

\[ \frac{\alpha_i}{1 + \alpha_i} (xP_i - P_{k, i}) > 2\Delta_i \quad (21) \]

and

\[ \frac{\alpha_i}{1 + \alpha_i} (xP_i - P_{k, i}) < 0 \quad (22) \]

since \( |\Delta_i| > |\Delta_i - b| \) if and only if \( 2\Delta_i < b < 0 \). The negativity constraint \( (22) \) holds if and only if \( xP_i < P_{k, i} \).

The lower bound \( (21) \) holds if and only if

\[ \alpha_i (xP_i - P_{k, i}) > 2\Delta_i (1 + \alpha_i) \quad (23) \]

Therefore \( (21) \) holds if and only if

\[ \alpha_i (xP_i - P_{k, i} - 2\Delta_i) > 2\Delta_i \quad (24) \]

If \( 2\Delta_i < xP_i - P_{k, i} \) then

\[ \alpha_i > \frac{2\Delta_i}{xP_i - P_{k, i} - 2\Delta_i} \quad (25) \]

and if \( 2\Delta_i > xP_i - P_{k, i} \) then

\[ \alpha_i < \frac{2\Delta_i}{xP_i - P_{k, i} - 2\Delta_i} \quad (26) \]

But if \( 2\Delta_i < xP_i - P_{k, i} \) then

\[ \alpha_i < \frac{2\Delta_i}{xP_i - P_{k, i} - 2\Delta_i} \quad (27) \]

or

\[ \alpha_i > \frac{2\Delta_i}{xP_i - P_{k, i} - 2\Delta_i} \quad (28) \]

Therefore if \( \alpha_i \in (-\Delta_i, 0) \) with

\[ A_i = -\frac{2\Delta_i}{xP_i - P_{k, i} - 2\Delta_i} > 0 \]

then \( (27) \) and \( (28) \) hold. So if \( A_i = \min_i \{A_i\} > 0 \) then the theorem holds for all states \( i \) that satisfy the inequality \( \Delta_i = (x - x^\infty) P_i < 0 \).

\[ \text{QED} \]
Algorithm 1 The Optimal Markov Chain Noise Benefit Algorithm

1: procedure MarkovChain($x_0$, $P$, $x^\infty$)  
2: $x_t \leftarrow x_0$  
3: repeat  
4: $x_t \leftarrow x_t \times P$  
5: $x_t \leftarrow$ NoisyStep($x_t$, $P$, $x^\infty$)  
6: until isConverged($x_t$)  
7: return $x_t$

8: procedure NoisyStep($x_t$, $P$, $x^\infty$)  
9: $n_t \leftarrow$ CalcNoise($x_t$, $n_{t-1}$, $x^\infty$)  
10: $\tilde{x}_t \leftarrow \frac{1}{\sum_n n_t} (x_t + n_t)$  
11: return $\tilde{x}_t$

12: procedure CalcNoise($x_t$, $P$, $x^\infty$)  
13: $\Delta \leftarrow (x_t - x^\infty) P$  
14: $L \leftarrow$ Length($\Delta$)  
15: $A \leftarrow [\Delta]_0$  
16: $k \leftarrow 0$

17: for $j \leftarrow 1, L$ do  
18: if $|\Delta[j]| > A$ then  
19: $A \leftarrow -\Delta[j]$  
20: $k \leftarrow j$

21: $n \leftarrow$ ZeroVector($L$)  
22: $n[k] \leftarrow$ Uniform(0, A)  
23: return $n$

Algorithm 2 The Blind Markov Chain Noise Benefit Algorithm

1: procedure MarkovChain($x_0$, $P$)  
2: $x_t \leftarrow x_0$  
3: repeat  
4: $x_t \leftarrow x_t \times P$  
5: $x_t \leftarrow$ NoisyStep($x_t$, $P$, $x_t$)  
6: until isConverged($x_t$)  
7: return $x_t$

8: procedure NoisyStep($x_t$, $P$, $x_t$)  
9: $n_t \leftarrow$ CalcNoise($x_t$, $n_{t-1}$, $x_t$)  
10: $\tilde{x}_t \leftarrow \frac{1}{\sum_n n_t} (x_t + n_t)$  
11: return $\tilde{x}_t$

12: procedure CalcNoise($x_t$, $n_{t-1}$, $x_t$)  
13: $\Delta \leftarrow x_t - n_{t-1} - x_t$  
14: $L \leftarrow$ Length($\Delta$)  
15: $A \leftarrow [\Delta]_0$  
16: $k \leftarrow 0$

17: for $j \leftarrow 1, L$ do  
18: if $|\Delta[j]| > A$ then  
19: $A \leftarrow -\Delta[j]$  
20: $k \leftarrow j$

21: $n \leftarrow$ ZeroVector($L$)  
22: $n[k] \leftarrow$ Uniform(0, A)  
23: return $n$

IV. MARKOV CHAIN EXPERIMENTAL RESULTS

The simulations below show that the proposed noise benefit applies to a wide range of Markov chain models. The three simulations show the evolution of the state density by direct computation of $x_{t+1} = x_t \times P$. Figures 1, 2, and 7 show the probability of several states over time. The first simulation applies noise to the 2-parameter Ehrenfest diffusion model. The simulation reaches steady state about 24% faster than the simulation without noise and provides evidence that the Markov Chain Noise Benefit theorem can apply to birth-death processes. The second simulation demonstrates that the Wright-Fisher population genetics model benefits from noise by decreasing the time to convergence. The third simulation shows that noise can speed simulations of a proposed chemical reaction whose state transition matrix derives from empirical measurement data.

A. Noise Benefits in the Ehrenfest Diffusion Model

The first simulation shows a noise benefit in the Ehrenfest diffusion model. Paul Ehrenfest proposed a diffusion model in the early 1900s as a statistical interpretation of the second law of thermodynamics [39, 40]. The model demonstrates the increase in entropy of a closed system over time [41]. The simulation shows that the noise benefit theorem
applies to a class of Markov models called birth-death processes. A birth-death process has the constraint $P_{ij} = 0$ if $|i - j| > 1$ [40, 42–45]. The simulation also demonstrates a noise benefit in a model that converges only in distribution. Figure 1 illustrates the noise benefit in an $N = 12$ molecule Ehrenfest diffusion simulation. Table I show how each state $i$ corresponds to a distribution of 12 molecules divided between two compartments $A$ and $B$.

**TABLE I: Number of molecules ($N = 12$) per compartment in simulation state $i$**

<table>
<thead>
<tr>
<th>State $i$</th>
<th>No. molecules in $A$</th>
<th>No. molecules in $B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

The simulation employed a 2-parameter generalized model with $s = 0.10$ and $t = 0.90$. The figure shows that the components approach their steady state values 24.2% faster on average with added noise (error < 0.5% of steady state).

The simplest Ehrenfest diffusion model uses a rectangular container with a permeable membrane separating two equally sized compartments called compartment $A$ and compartment $B$ [46–48]. The container holds $N$ gas molecules that the membrane allows to pass between compartments (Figure 3).

The model randomly selects a molecule at each time step and then moves that molecule to the other compartment. $x(t)$ denotes the number of molecules in compartment $A$ at each time step. So $x(t) \in \{0, 1, 2, \ldots, N\}$. The simulation tends toward a steady state distribution with maximal entropy as $t \to \infty$ [49]. The Ehrenfest model is a birth-death process because $x(t)$ either increments or decrements by one at each time step [50]. Suppose the container contains $N$ molecules and has $0 < M < N$ molecules in compartment $A$ at time $t$. Then

$$x(t) = M$$

and

$$P[x(t + 1) = M - 1] = \frac{M}{N}$$

$$P[x(t + 1) = M + 1] = 1 - \frac{M}{N}$$

The Markov chain $x(t)$ evolves according to the state transition matrix $P$ where

$$P_{ij} = \begin{cases} \frac{N}{s} & j = i + 1 \\ \frac{1}{t} & j = i - 1 \\ \frac{1}{M} [(1-s)N + i(s-t)] & j = i \\ 0 & \text{else} \end{cases}$$

for $s, t \in [0, 1]$ and integers $0 \leq i, j \leq N$ [53]. The Krafft-Schaefer model converges with probability one only for the trivial case where one of the compartments is a perfect sink (when $s = 0$ or $t = 0$). The model weakens to convergence in distribution for all other $s$ and $t$ such that $0 < s, t < 1$.

Figure 1 shows a simulation that initialized $x_0$ as a normalized random state vector. This represents starting the diffusion simulation with uncertainty in the system’s configuration. The simulation used $s = 0.10$ and $t = 0.90$. 

**FIG. 3:** (Color online) Two compartment Ehrenfest diffusion model. The figure illustrates the diffusion experiment of Ehrenfest. The box contains $N = 20$ molecules. The compartments $A$ and $B$ partition the box. $x(t)$ represents the number of molecules in box $A$ at time $t$. This example assumes $x(t) = 16$. The simulation randomly selects a molecule at each time step (red circle) and moves the selected molecule to the other compartment (red arrow). Here $x(t + 1) = 16 - 1 = 15$ since one molecule moves from $A$ to $B$. The model exhibits a dynamic equilibrium because molecules continue to shuttle across the membrane for all $t$. In particular: $x(t) \neq x(t + 1)$. So the occupancy $x(t)$ converges in distribution.
Consider a gene with 2 alleles (A1 and A2) in a population with N diploid individuals. The population contains 2N copies of the gene since each diploid individual has 2 copies of the gene. Let the state vector x(t) represent the allele distribution at time t [58]. Then at time t:

\[
x_0(t) = P[0 \text{ copies } A_1, 2N \text{ copies } A_2]
\]
\[
x_1(t) = P[1 \text{ copies } A_1, 2N - 1 \text{ copies } A_2]
\]
\[
x_2(t) = P[2 \text{ copies } A_1, 2N - 2 \text{ copies } A_2]
\]
\[
\vdots
\]
\[
x_{2N}(t) = P[2N \text{ copies } A_1, 0 \text{ copies } A_2].
\]

The Wright-Fisher model produces successive generations with a 2-step process (Figure 5). The model first creates N pairs of parents selected randomly and with replacement from the population. Then each pair produces a single offspring with its genotype inherited by selecting one gene from each parent. All parents die after mating.

The allele distribution x(t) is a Markov chain that advances by random sampling with replacement from the pool of parent genes (Figure 6) [59, 60]. The density of alleles evolves according to a binomial probability density with

\[
P[x(t + 1) = j \mid x(t) = i] \sim \text{Bin} \left(j; 2N, \frac{i}{2N} \right)
\] (29)

Thus the Markov chain transition matrix has elements [58]:

\[
P_{i,j} = \binom{2N}{j} \left( \frac{i}{2N} \right)^j \left( 1 - \frac{i}{2N} \right)^{2N-j}.
\]

Figure 6 also demonstrates how the allele distribution x(t) converges to the steady-state. x(t) converges with probability one to either of the homozygous populations - either (A1, A1) or (A2, A2) [61]. This convergence is much stronger than the convergence in distribution found in the Ehrenfest diffusion model.

The Wright-Fisher simulation used a population N = 50 diploid individuals. The simulation tracked the allele distribution of a diallelic gene: A1 and A2. It initialized

FIG. 4: (Color online) Two compartment Krafft-Schaefer asymmetric diffusion model. The figure illustrates the membrane “preference” in the asymmetric Krafft-Schaefer diffusion model. Here s ≫ t. So P[B → A] ≫ P[A → B] for a particular molecule (indicated by relative size of arrows). The asymmetry shifts the equilibrium to the left so that more molecules tend to accumulate in A at steady state.

FIG. 5: (Color online) Wright-Fisher mating. The figure illustrates how the Wright-Fisher model produces successive generations. Each doublet in the first row represents the genotype of a diploid individual from a population N = 4. Each organism possesses a pair of alleles (blue = A1 and red = A2). The two middle rows show how the model randomly pairs individuals with replacement to form mates. Each doublet in the last row represents an offspring. The offspring inherit one allele (A1 or A2) randomly from each parent. Then the simulation “kills” the t = n population and the offspring become the new t = n + 1 population.

B. Noise Benefits in a Population Genetics Model

The second simulation shows a noise benefit in the Wright-Fisher population genetics model. The Wright-Fisher model uses a Markov chain to simulate stochastic genotypic drift during successive generations [54–56]. Figure 2 illustrates the noise benefit in a simulation with 2 alleles and N = 50 diploid individuals.

The Wright-Fisher model applies generally to populations under the following assumptions [57]:

1. the population size N remains constant between generations

2. no selective difference between alleles

3. non-overlapping generations.

Consider a gene with 2 alleles (A1 and A2) in a population with N diploid individuals. The population contains 2N copies of the gene since each diploid individual has 2 copies of the gene. Let the state vector x(t) represent

to slow convergence and highlight the noise benefit. The asymmetry due to s = 0.10 and t = 0.90 collapses the dominant eigengap |λ1| − |λ2| where λi is the ith largest magnitude eigenvalue. This increases the time for the simulation to reach steady state. A similar benefit exists for all s and t in (0, 1). A wider eigengap |λ1| − |λ2| ensures that the chain quickly converges toward steady state. This results in a smaller noise benefit.

FIG. 4: (Color online) Two compartment Krafft-Schaefer asymmetric diffusion model. The figure illustrates the membrane “preference” in the asymmetric Krafft-Schaefer diffusion model. Here s ≫ t. So P[B → A] ≫ P[A → B] for a particular molecule (indicated by relative size of arrows). The asymmetry shifts the equilibrium to the left so that more molecules tend to accumulate in A at steady state.
FIG. 6: (Color online) Markov dynamics of a Wright-Fisher genotype. Each of the 6 circles for $t = 0$ represents an allele for a particular gene (blue = $A_1$ and red = $A_2$). The Wright-Fisher model generates the $t = 1$ offspring by randomly sampling the $t = 0$ population with replacement. The connections indicate the surviving genes and their offspring. The $A_1$ allele becomes extinct by the fourth generation in this example. The steady state for this example is homozygous ($A_2, A_2$) because future generations can no longer inherit the extinct $A_1$ gene.
aggregate NMR data using weighted least squares:

\[ P = \begin{bmatrix}
0.9274 & 0.0700 & 0.0025 & 8 \times 10^{-5} & 10^{-5} & 10^{-5} \\
0.0500 & 0.8395 & 0.1000 & 0.0100 & 0.0004 & 0.0001 \\
0.0600 & 0.0600 & 0.8495 & 0.0300 & 0.0004 & 0.0001 \\
0.0500 & 0.0100 & 0.0400 & 0.5400 & 0.0600 & 0.3000 \\
0.0500 & 0.0200 & 0.0200 & 0.0500 & 0.8595 & 0.0005 \\
0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.99953
\end{bmatrix} \]

corresponding to steady-state probability density \( x^\infty \):

\[ x^\infty = \begin{bmatrix}
0.026 & 0.017 & 0.013 & 0.002 & 0.002 & 0.942
\end{bmatrix}. \]

Figure 8 summarizes the principal reaction pathways.

We used the experimental \( ^{29} \text{Si} \) NMR data reported earlier [62] to initialize the species concentrations to

\[ x_0 = \begin{bmatrix}
0.430 & 0.260 & 0.220 & 0.060 & 0.030 & 0.000
\end{bmatrix} \]

and advanced the Markov chain to simulate the crystallization of zeolite.

Our simulations show that noise benefits the empirical estimation but the observed benefit was small. The performance metric showed a strong benefit of 18.1% despite some states experiencing only minimal noise benefits (Figure 7.a and c). This is because the noise quickly moved a few components to their steady-state value (Figure 7.b). The Markov Chain Noise Benefit Theorem could not provide additional benefit to the system after this initial boost because the theorem relies on the magnitude of the component closest to its steady state value. Several components converged within a few time-steps.

So the theorem-based noise added only small corrections to the states for the rest of the simulation. This shows that the theorem confers a larger benefit to systems with states that converge at approximately the same rate. But other Markov systems still receive some noise benefit.
V. MARKOV CHAIN NOISE BENEFIT THEOREM SIMULATION

Two final simulations show the noise benefit that exists for Markov chain simulations. The simulations show how the Markov Chain Noise Benefit Theorem might speed convergence in modern algorithms such as the Google PageRank™ link analysis algorithm [96–98]. The PageRank algorithm constructs a probability density that represents the likelihood that a person randomly clicking on links will arrive at a particular page over all indexed pages on the Internet [99]. The algorithm operates on a dataset called the Google matrix. This matrix is equivalent to a Markov state transition matrix spanning tens of billions of dimensions [100, 101]. The Noise Benefit Theorem shows that the algorithm should benefit from noise.

A. One-step Markov Chain Simulation

The first simulation shows that a Markov chain can benefit from additive noise (Figure 9). The simulation shows the benefit after one time-step as a decrease of the absolute error between the posterior state density and the stationary state density. Table II and Table III show a large decrease in the absolute error in the noisy simulations compared to the no-noise simulations.

Table II summarizes the one-step experiment with and without noise. The simulation classified the states as satisfying either the conditions of the Markov Chain Noise Benefit Theorem $\Delta_i = (x - x^\infty) P_i > 0$ or the Corollary $\Delta_i = (x - x^\infty) P_i < 0$. It set the noise strength to $A = \min \{a_i\}$ for each class in accord with the theorem. This gave $A = 0.0682$ for the states with $\Delta_i > 0$ and $A = -0.1594$ for the states with $\Delta_i < 0$. The simulation calculated the total absolute error for each class using the respective values for the noise strength $A$.

<table>
<thead>
<tr>
<th>States satisfying MC Theorem</th>
<th>States satisfying corollary</th>
</tr>
</thead>
<tbody>
<tr>
<td>no noise</td>
<td>0.1547</td>
</tr>
<tr>
<td>with noise</td>
<td>0.0547</td>
</tr>
<tr>
<td>error decrease (%)</td>
<td>64.64%</td>
</tr>
<tr>
<td></td>
<td>53.20%</td>
</tr>
</tbody>
</table>

Table III summarizes a simulation with and without noise that does not have access to signs of the inequality. It shows that a noise benefit exists even if the simulation cannot classify individual states according to $\Delta_i > 0$ or $\Delta_i < 0$. The table summarizes the relative improvement over all $N = 6$ states when setting the noise strength to $A = \text{sign}(a_i) \cdot \min(|a_i|)$. This gave $A = \min(0.0682, 0.1594) = 0.0682$. The simulation calculated the total absolute error for the posterior state density using this value of $A$.

<table>
<thead>
<tr>
<th>Noise benefits in one-step Markov chain simulation – unknown error sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>no noise</td>
</tr>
<tr>
<td>with noise</td>
</tr>
<tr>
<td>error decrease (%)</td>
</tr>
</tbody>
</table>

The Markov Chain Noise Benefit Theorem ensures that there exists a noise distribution that reduces the state error. Figure 9 illustrates this because it shows that the error decreases as the noise strength increases from zero. The theorem and corollary also establish that past some noise strength ($A > 0$) the error will increase. Thus properly signed noise with magnitude less than $A$ guarantees that the absolute error will be lower in the noisy simulation than in the no-noise simulation.

Figure 9 shows an example where three of the $N = 6$ states obey the inequality (13) in the main theorem and the remaining three states obey the inequality (18) in the corollary. Not all transition matrices $P$ have this even splitting. But any given matrix will have at least one state that satisfies each case since the sum of the signed errors must equal 0.

The simulation generated a Markov chain from a fixed random transition matrix where $\hat{P}_{i,j} = P[X_{k+1} = j|X_k = i] \sim U(0,1)$. The simulation used $N = 6$ states. The theorem and corollary guarantee the benefit for transition matrices with any finite dimension. But uniformly chosen transition matrices tend to generate a uniform stationary density:

$$x_j^\infty = \frac{1}{N}. \quad (32)$$

We transformed each transition probability by $U(0,1) + \epsilon$ to construct a network of states with non-uniform importance. We chose $\epsilon > 0.04$ to avoid numerical instability. This gives a transition matrix

$$\hat{P}_{i,j} = \frac{U_{i,j}}{V_{i,j} + \epsilon}. \quad (33)$$

where $U_{i,j} \sim U(0,1)$ and $V_{i,j} \sim U(0,1)$. We normalized the rows of $\hat{P}$ to form a proper stochastic matrix:

$$P_{i,j} = \frac{\hat{P}_{i,j}}{\sum_{k=1}^{N} \hat{P}_{i,k}}. \quad (34)$$

We chose the initial state density $x$ as the uninformed prior $\text{[36]}$ (uniform distribution) over the 6 states so that

$$x_j = \frac{1}{N}. \quad (35)$$

We used Matlab R2009b to perform the simulations.
FIG. 9: (Color online) Noise benefits in Markov chain density estimation. These figures show the relation between the error magnitude of each Markov state and the noise strength \( a_i \). The simulation used a 6-state Markov chain and the figure shows the single-step absolute errors by state. Each of the six states satisfied either (a) the Markov Chain Noise Benefit Theorem: \( (x - x^\infty) P_i > 0 \) or (b) the Corollary: \( (x - x^\infty) P_i < 0 \). (a) Three states satisfy the inequality \( (x - x^\infty) P_i > 0 \) in this simulation. Each curve represents the absolute error \( \|\hat{P} - x^\infty\| \) of the \( i^{th} \) state as \( a_i \) increases. The standard zero-noise condition corresponds to \( a_i = 0 \). Each state has an optimal noise level \( A_i \) indicated by the point where the curve meets the \( a_i \)-axis. The optimal noise \( A_i \) will exactly drive the state to its stationary value. The Markov Chain Noise Benefit Theorem first shows that the benefit exists for all \( a_i < A_i \). The theorem also guarantees the existence of a global \( A = \min \{ A_i \} > 0 \) such that any noise \( a < A \) benefits every state that satisfies the inequality. All curves decrease (strictly) monotonically until they reach \( A_i \). Thus any point between the no-noise condition and \( A_i \) shows some benefit and \( A = \min \{ A_i \} \) satisfies this constraint for each such state. (b) Three states satisfy the alternative inequality \( (x - x^\infty) P_i < 0 \). These correspond to the states that satisfy the corollary. The corollary ensures a point \( A \) so that any noise-strength less than \( A \) benefits every such state.

with transition matrix \( P \):

\[
P = \begin{bmatrix}
0.038 & 0.040 & 0.077 & 0.070 & 0.065 & 0.710 \\
0.017 & 0.109 & 0.140 & 0.128 & 0.234 & 0.372 \\
0.014 & 0.022 & 0.062 & 0.174 & 0.005 & 0.723 \\
0.027 & 0.053 & 0.068 & 0.184 & 0.058 & 0.611 \\
0.071 & 0.075 & 0.115 & 0.132 & 0.011 & 0.696 \\
0.181 & 0.177 & 0.484 & 0.017 & 0.068 & 0.073
\end{bmatrix}
\]

corresponding to steady-state probability density \( x^\infty \):

\[
x^\infty = \begin{bmatrix}
0.089 & 0.102 & 0.241 & 0.094 & 0.065 & 0.408
\end{bmatrix}
\]

B. Two-step Markov Chain Simulation

The second simulation shows that the noise benefits in the one-step simulation extend over successive time steps (Figure 10). We measure the benefit as a decrease in the absolute error between the posterior state density and the stationary state density. The simulation also shows that even suboptimal noise in one time step can still benefit successive steps. The proof guarantees that there exists a noise density that will reduce the error over multiple time steps.

We generated a transition matrix \( P \) using the same procedure as in the one-step simulation (34):

\[
P = \begin{bmatrix}
0.147 & 0.013 & 0.051 & 0.667 & 0.062 & 0.061 \\
0.158 & 0.030 & 0.088 & 0.622 & 0.012 & 0.090 \\
0.078 & 0.061 & 0.095 & 0.582 & 0.077 & 0.108 \\
0.138 & 0.106 & 0.055 & 0.565 & 0.039 & 0.098 \\
0.171 & 0.085 & 0.213 & 0.085 & 0.170 & 0.276 \\
0.048 & 0.028 & 0.070 & 0.804 & 0.030 & 0.020
\end{bmatrix}
\]

This corresponds to the steady-state probability density \( x^\infty \):

\[
x^\infty = \begin{bmatrix}
0.129 & 0.077 & 0.068 & 0.582 & 0.048 & 0.094
\end{bmatrix}
\]

VI. CONCLUSION

We have shown that noise can benefit Markov chain estimation by speeding up the convergence time if the algorithm can calculate the sign of the state error. We have also shown how a simulation can use estimates of the error magnitude to update its current estimate of the underlying state density. Simulations confirm that noise can benefit a single-step or multi-step system even if the system has insufficient information to determine the optimal noise.
FIG. 10: (Color online) Multi-cycle noise benefits in Markov chain density estimation. This figure shows that the noise benefits apply for successive Markov steps. It further shows that even suboptimal noise in one iteration can still benefit successive steps. The simulation evaluated the deciding inequalities (13) and (18) for a single state at two successive time steps and used only the sign (+ or −) to determine the direction of beneficial noise for the state. The plot shows the relation between the state’s absolute error and the noise magnitude during the “first step” and “second step” (with the appropriate sign). The origin of the $a_{1}^{(\text{step 1})}$ and $a_{2}^{(\text{step 2})}$ axes corresponds to a zero-noise 2-step Markov chain. The optimal noise corresponds to $a_{1}^{(\text{step 1})} \approx 0.165$ during the first step. Then there is a strictly positive value for $a_{2}^{(\text{step 2})}$ that yields a lower error than the zero-noise case even if the system applies suboptimal noise during the first step — such as $a_{1} = 0.10$.

Versions of the Markov Chain Noise Benefit Theorem may well hold for weaker assumptions and other Markov chain models. An open question is whether the results hold for noise-perturbed Markov transition matrices instead of noise-perturbed state densities. This may apply to simulations with noisy estimates of the transition matrix or to simulations with transition matrix estimates based on only a few observations. This might also apply to specific MCMC algorithms under suitable assumptions. Adaptive algorithms may be able to find optimal noise amounts in many of these cases.

[19] M. Guerriero, S. Marano, V. Matta, and P. Willett,
Suppose there exists some \( Y \) for all states \( i \) for all \( n \in \mathbb{N} \).

Appendix A: Proof of the Markov Chain Noise Benefit Theorem

Markov Chain Noise Benefit Theorem. Suppose \( M \) is a finite time-homogeneous Markov chain with \( N \) states and transition matrix \( P \). Suppose further that \( M \) is irreducible and aperiodic. Then for all non-stationary state density vectors \( x \) there exists a noise benefit in the sense that there exists some \( A > 0 \) so that for all \( a \in (0, A) \):

\[
||\tilde{x}P - x^\infty||_1 < ||xP - x^\infty||_1
\]

for all states \( i \) with

\[
\Delta_i = (x - x^\infty)P_i > 0
\]

where

\[
\tilde{x} = \frac{1}{1 + a} (x + n)
\]

is the normalized state vector after adding a noise vector \( n \) with only one non-zero component:

\[
n_j = \begin{cases} a & j = k \\ 0 & j \neq k \end{cases}
\]

for any \( k \) that satisfies

\[
\Delta_k = (x - x^\infty)P_k > 0.
\]

Proof. Fix \( x \) as a state vector of the Markov chain \( M \). Note first that \( \tilde{x} \) is a probability density function over the states of \( M \) because of (a) and (b) below:

(a). \( \tilde{x} \) is a \( N \)-vector with \( \tilde{x}_i \geq 0 \) since

\[
[\tilde{x}]_i = \left[ \frac{1}{1 + a} (x + n) \right]_i = \left[ \frac{1}{1 + a} x_i \right]_i 
\]

for \( i \neq k \) and

\[
\geq \frac{1}{\max(1, A)} x_i \geq 0 \]

since \( a > 0 \) and \( A > 0 \).

(b). \( \sum \tilde{x}_i = 1 \) since

\[
\sum_{i=1}^{N} \tilde{x}_i = \frac{1}{1 + a} \left( \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} n_i \right) = \frac{1}{1 + a} (1 + a) = 1.
\]

Note that

\[
||\tilde{x}P - x^\infty||_1 < ||xP - x^\infty||_1 \]

The proof proceeds by showing that such an \( A_i \) exists for each component \( i \) that satisfies \( \Delta_i = (x - x^\infty)P_i > 0 \). This will complete the proof because \( (0, A) = \cap_{i=1}^{N} (0, A_i) \neq 0 \) since \( N < \infty \) and \( \mu ([0, A_i]) > 0 \) for each \( A_i \).

Let \( i \) in \( 1 \leq i \leq N \) be any state that satisfies the inequality \( \Delta_i = (x - x^\infty)P_i > 0 \). Choose \( k \) in \( 1 \leq k \leq N \) and define

\[
\tilde{x} = \frac{1}{1 + a_i} (x + n)
\]
with

\[ n_j = \begin{cases} a_i & j = k \\ 0 & j \neq k \end{cases} \quad (A18) \]

and \( a_i > 0 \). Then

\[
|\mathbf{x}_P - x^\infty|_i = |(\mathbf{x}_P)_i - x^\infty|_i \\
= |(\mathbf{x}_P)_i - x^\infty P_i|_i
\]

(A19)

and

\[
|\mathbf{x}_P - x^\infty|_i = \left| \sum_{j=1}^{n} x_j P_{j,i} - x_j^\infty P_{j,i} \right|
\]

(A21)

\[
= \left| \sum_{j=1}^{n} \frac{1}{1 + a_i} (x + n) P_{j,i} - x_j^\infty P_{j,i} \right|
\]

(A22)

\[
= \left| \sum_{j=1}^{n} \frac{1}{1 + a_i} (x_j + n_j) P_{j,i} - x_j^\infty P_{j,i} \right|
\]

(A23)

\[
= \left| \frac{1}{1 + a_i} \sum_{j=1}^{n} (x_j P_{j,i} + n_j P_{j,i}) - x_j^\infty P_{j,i} \right|
\]

(A24)

\[
= \frac{1}{1 + a_i} \left( \sum_{j=1}^{n} x_j P_{j,i} + n_j P_{j,i} \right) - x_j^\infty P_{j,i}
\]

(A25)

\[
= \frac{1}{1 + a_i} \left( \sum_{j=1}^{n} x_j P_{j,i} + n_j P_{j,i} \right) - x_j^\infty P_{j,i}
\]

(A26)

\[
= \left| \sum_{j=1}^{n} x_j P_{j,i} + n_j P_{j,i} \right|
\]

(A27)

so

\[
|\mathbf{x}_P - x^\infty|_i = \left| \Delta_i - \frac{a_i}{1 + a_i} (x P_i - P_{k,i}) \right|
\]

(A31)

Now \( \Delta_i > 0 \) by hypothesis. Thus

\[
\left| \Delta_i - \frac{a_i}{1 + a_i} (x P_i - P_{k,i}) \right| < |\Delta_i| \quad (A32)
\]

if and only if

\[
\frac{a_i}{1 + a_i} (x P_i - P_{k,i}) > 0 \quad (A33)
\]

and

\[
\frac{a_i}{1 + a_i} (x P_i - P_{k,i}) < 2\Delta_i \quad (A34)
\]
since $|\Delta_i| > |\Delta_i - b|$ if and only if $0 < b < 2\Delta_i$. The positivity constraint (A33) holds if and only if $xP_i > P_{k,i}$. The upper bound (A34) holds if and only if

$$a_i (xP_i - P_{k,i}) < 2\Delta_i (1 + a_i).$$

Therefore (A34) holds if and only if

$$a_i (xP_i - P_{k,i} - 2\Delta_i) < 2\Delta_i.$$  (A35)

If $2\Delta_i < xP_i - P_{k,i}$ then

$$a_i < \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i}.$$  (A37)

and if $2\Delta_i > xP_i - P_{k,i}$ then

$$a_i > \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i}.$$  (A38)

But if $2\Delta_i > xP_i - P_{k,i}$ then $\frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} < 0$. So any $a_i > 0$ suffices. Thus either

$$a_i > 0$$

if $2\Delta_i > xP_i - P_{k,i}$ (A39)

or

$$a_i < \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i}$$

if $2\Delta_i < xP_i - P_{k,i}$. (A40)

Therefore if $a_i \in (0, A_i)$ with $A_i = \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} > 0$ then (A39) and (A40) hold. So if $A = \min_i \{A_i\} > 0$ then the theorem holds for all states $i$ that satisfy the inequality $\Delta_i = (x - x^\infty) P_i > 0$.

QED