A Maximum Entropy Model for Flocking in 2D

Russell Burton

Abstract

A maximum entropy model originally constructed in 3D to describe ordered flocks of real birds is repurposed to describe an ordered flock in 2D. The 2D flock obeys Vicsek dynamics, and parameters of the Vicsek model are varied against the inferred interaction parameters of the maximum entropy model. The inferred interaction parameters scale as anticipated, although the data is quite noisy. Directions for further research are suggested.

This paper represents my own work in accordance with University regulations.

1 Introduction and Motivations

Quantitatively describing collective motion in flocks of animals has long remained a challenge [1]. Much like in certain problems of statistical physics, flocks contain many seemingly identical agents that decide on a spontaneous direction of motion, despite apparent randomness in motion of individual agents. To a physicist, this type of problem may suggest an analogy to the theory of ferromagnets. Given resemblances of this sort, many have applied tools of statistical mechanics to describe flocking: defining order parameters, identifying phase transitions, computing scaling exponents, and the like [2-6].

Of the many tools native to statistical mechanics, entropy has found success in fields besides physics. The discipline of information theory serves as a prime example. Claude Shannon's seminal work in information theory defined a notion of "information" in accordance with a set of axioms in accordance with intuitive notions of information [7]. At the heart of the theory lies entropy. For a given probability distribution, $P(\vec{x})$, its Shannon entropy is

$$S[P] = -\sum_{\mathbf{x}} P(\mathbf{x}) \ln P(\mathbf{x}). \tag{1.1}$$

Intuitively Shannon entropy is a measure of the amount of "randomness" a distribution describes: the higher entropy, the more random the distribution. In a similar spirit, Shannon called a decrease in information entropy an increase in information. For example in describing a coin flip with outcomes,

H and T, the distribution with the least information (maximum entropy) is a 50:50 distribution $P(H) = P(T) = 1/2$ with entropy $\ln 2$. A coin that always flipped heads would have the least entropy for such a configuration space, since we have perfect information about the outcome of the coin flip.

E.T. Jaynes is credited for being the first to take seriously the significance of distributions that are maxima of Shannon entropy in the context of model building [8]. He demonstrated an inference procedure to construct maximum entropy (MaxEnt) distributions given some state variables $\vec{x}$, observables $f_\mu(\vec{x})$, and averages from experiment $\langle f_\mu \rangle_{exp}$. Borrowing the intuition from Shannon, he reasoned that maxima of Shannon entropy contain the least amount of presupposed information about the nature of the system.

More recently principles of maximum entropy have been used to recast biological questions into statistical mechanics problems [3] [4] [9] [10] [11]. In particular Bialek and collaborators analyzed maximum entropy models (MEMs) describing real flocks of birds [3] [4]. The MEM used in the first paper used was mathematically identical to a 3D classical Heisenberg spin model in equilibrium. The 3D Heisenberg model orders at low temperature, just like the ferromagnetic system it describes [12]. Since the flock of real birds modeled was experimentally known to magnetize, so the choice of model was appropriate in that regard.

Using the same MEM adapted for 2D, however, results in an MEM equivalent to an XY spin system in 2D, which does not possess an ordered phase [13]. The goal of this thesis is to apply a 2D version of the local correlation function used in Bialek's MEM to a 2D ordered flock. The primary aim is to check intuitively how the MEM's interaction parameters scale with the real flock's parameters.

2 Fitting an MEM to Experimental Data

First let us introduce Jaynes' method of finding MaxEnt distributions. Start with a system described by real state variables $\mathbf{x} = \{x_1, \ldots, x_n\}$ and K real-valued observables $f_\mu(\mathbf{x})$, each with averages measured from experiment, $\langle f_\mu(\mathbf{x})\rangle_{exp}$.

As described in section 1, Shannon entropy is defined as

$$S[P] = -\sum_{\mathbf{x}} P(\mathbf{x}) \ln P(\mathbf{x}). \tag{2.1}$$

The goal is to find the distribution P^* that maximizes this functional subject to the constraint that it also reproduces the experimental data (i.e. $\langle f_\mu(\mathbf{x})\rangle_{P^*} = \langle f_\mu(\mathbf{x})\rangle_{exp}$). For proof of the uniqueness of P^* and that it is indeed a maximum of entropy, not a minimum, refer to [14].

We compute P^* via the technique of Lagrange multipliers. The normalization of the distribution may be conveniently enforced by including an additional $f_0(\mathbf{x}) = 1$ function to the set of observables. The Lagrange multiplier equation is

$$S[P, \{\lambda_\mu\}] = S[P] - \sum_{\mu=0}^{K} \lambda_\mu \Big[\langle f_\mu(\mathbf{x})\rangle_P - \langle f_\mu(\mathbf{x})\rangle_{exp} \Big] \tag{2.2}$$

where $P^*(\mathbf{x})$ maximizes this quantity with respect to both P and the λ_μ.

Taking first the functional derivative over P,

$$\begin{aligned}\frac{\delta S[P,\{\lambda_\mu\}]}{\delta P} &= \frac{\delta}{\delta P(\mathbf{x})}\left(-\int P(\mathbf{x}')\ln P(\mathbf{x}')d\mathbf{x}' - \sum_{\mu=0}^{K}\lambda_\mu\int f_\mu(\mathbf{x}')P(\mathbf{x}')d\mathbf{x}'\right)\\ &= -\ln P(\mathbf{x}) - 1 - \sum_{\mu=0}^{K}\lambda_\mu f_\mu(\mathbf{x}).\end{aligned} \tag{2.3}$$

After setting the previous equal to zero and rearranging the terms,

$$P^*(\mathbf{x}) = \exp\left[-1 - \lambda_0 f_0(\mathbf{x}) - \sum_{\mu=1}^{K}\lambda_\mu f_\mu(\mathbf{x})\right]. \tag{2.4}$$

We may substitute $f_0(\mathbf{x}) = 1$, and after doing so P has the simple form

$$P^*(\mathbf{x}) = \frac{1}{Z}\exp\left[-\sum_{\mu=1}^{K}\lambda_\mu f_\mu(\mathbf{x})\right] \tag{2.5}$$

where $Z(\{\lambda_\nu\} \equiv \exp[1+\lambda_0]$. The maximization of $S[P,\{\lambda_\mu\}]$with respect to the λ_μ after inserting $P = P^*$ enforces the constraints $\langle f_\mu\rangle_{exp} = \langle f_\mu\rangle_{P^*}$, or

$$\langle f_\mu(\mathbf{x})\rangle_{exp} = \frac{1}{Z(\{\lambda_\mu\})}\sum_{\mathbf{x}} f_\mu(\mathbf{x})\exp\left[-\sum_{\mu=1}^{K}\lambda_\mu f_\mu(\mathbf{x})\right]. \tag{2.6}$$

We enforce the normalization constraint by setting $\langle f_0(\mathbf{x})\rangle_P = \int d\mathbf{x}P(\mathbf{x}) = \langle f_0(\mathbf{x})\rangle_{exp} = 1$.

For the remaining λ_μ we can use a trick familiar to statistical physics, ex-

ploiting the partition function

$$\frac{\partial}{\partial \lambda_\mu} \ln Z(\{\lambda_\mu\}) = \frac{1}{Z}\frac{\partial Z}{\partial \lambda_\mu} = \frac{1}{Z}\int f_\mu \exp\left(\sum_\mu \lambda_\mu f_\mu\right) d\mathbf{x} = \langle f_\mu \rangle_P \qquad (2.7)$$

Thus if we can compute Z as a function of λ_μ, we can use the constraint equations $\frac{\partial}{\partial \lambda_\mu} = \langle f_\mu \rangle_{exp}$ to solve for the λ_μ.

This is a good opportunity to explain more the intuition about MEMs. Plugging the formula for P back into the Lagrange multiplier form of S shows that

$$\begin{aligned}
S[P^*, \{\lambda_\mu\}] &= -\int d\mathbf{x} P^*(x) \ln P^*(x) - \sum_{\mu=0}^{K} \lambda_\mu \Big[\langle f_\mu(\mathbf{x})\rangle_{P^*} - \langle f_\mu(\mathbf{x})\rangle_{exp}\Big] \\
&= -\int d\mathbf{x} P^*(x) \left(-\ln Z + \left[-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x})\right]\right) - \sum_{\mu=0}^{K} \lambda_\mu \Big[\langle f_\mu(\mathbf{x})\rangle_{P^*} - \langle f_\mu(\mathbf{x})\rangle_{exp}\Big] \\
&= \int d\mathbf{x} \ln Z P^*(\mathbf{x}) + \sum_\mu \lambda_\mu \langle f_\mu \rangle_{P^*} - \sum_{\mu=0}^{K} \lambda_\mu \Big[\langle f_\mu(\mathbf{x})\rangle_{P^*} - \langle f_\mu(\mathbf{x})\rangle_{exp}\Big] \\
&= \ln Z + \sum_{\mu=0}^{K} \lambda_\mu \langle f_\mu(\mathbf{x})\rangle_{exp} = \ln Z + \exp\left[\sum_\mu \lambda_\mu \langle f_\mu(\mathbf{x})\rangle_{exp}\right] \\
&= -\ln P^*(\{f_\mu = \langle f_\mu \rangle_{exp}\}) = -\langle \ln P^* \rangle_{exp} \equiv -\mathcal{L}(\{\lambda_\mu\}).
\end{aligned} \qquad (2.8)$$

The final line describes the minus log probability of P^* reproducing the experimental averages. The quantity $\mathcal{L}$ will later be maximized by adjusting the interaction range of the MEM, and the above derivation provides some

intuition about what exactly $\mathcal{L}$ is. In particular, maximizing $\mathcal{L}$ with respect to interaction length will *minimize* the entropy, because it is choosing the interaction range that maximizes the odds that $f_\mu = \langle f_\mu \rangle_{exp}$.

3 The Vicsek Model

In this paper the experimental system in question is a flock with Vicsek dynamics. The Vicsek model describes a simple dynamics for a collection of particles, nonetheless exhibiting collective motion in certain parameter regimes [2].

The Viscek model describes N particles in a 2D square with periodic boundary conditions. At a given time t, each particle has position $\vec{r_i}(t) = (x_i(t), y_i(t))$ and instantaneous direction of motion $\vec{s_i}(t) = (s_i^x(t), s_i^y(t))$, where $||\vec{s_i}|| = 1$. Each particle moves with the same constant speed ν. The dynamics are executed in discrete time steps, and for our purposes the time step will be set to $\Delta t = 1$. The equations of motion for each θ are

$$\theta_i(t+1) = \langle \theta_j(t) \rangle_{m^i} + \eta_i(t) \tag{3.1}$$

where $\langle \cdots \rangle_{m^i}$ is an average of the θ_j over the $\{m^i\}$ set of neighbors for i, and η_i are noise terms. To be specific, the average angle of neighbors $\langle \theta_j(t) \rangle_{m^i}$ is defined as $\arctan \frac{\sum_j m_{ij} \sin(\theta_j)}{\sum_j m_{ij} \cos(\theta_j)}$ in this context, with m_{ij} as the adjacency matrix that defines the neighborhood sets. As for the noise term, this paper follows Vicsek's convention to pick η_k as independent, identically distributed random variables drawn from the uniform distribution $[-\eta/2, \eta/2]$.

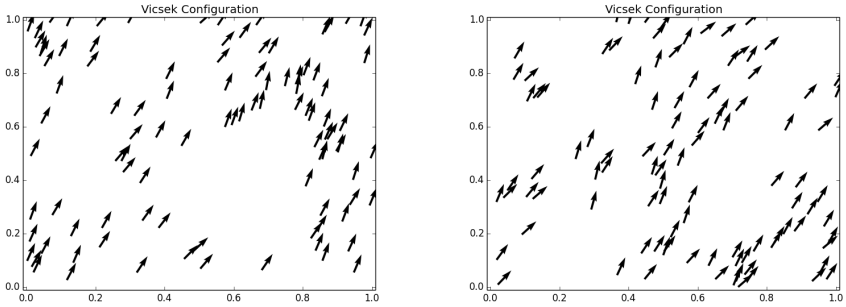

Figure 1: Snapshots of Vicsek flocks typical of what were probed in this paper. In the first image the noise $\eta = 0.3$ and in the second image $\eta = 0.5$.

The particles update their positions each time step according to the equation

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \nu \vec{s}_i(t). \tag{3.2}$$

Unlike Vicsek's original paper, the flocks simulated in this paper will have topological, not metric, interactions. Thus the length of the box, L, is irrelevant to the Vicsek dynamics, so we choose $L = 1$. Since we also chose $\Delta t = 1$, the velocity ν simply measures the length each particle moves each frame, measured in units of box length.

4 Constructing the MEM from local correlations

As shown in section 2, given observables $f_\mu(\mathbf{x})$ the MEM take the form

$$P = \frac{1}{Z} \exp\left(- \sum_\mu \lambda_\mu f_\mu(\mathbf{x}) \right) \tag{4.1}$$

with the λ_μ chosen via the constraint $\langle f_\mu \rangle_P = \langle f_\mu \rangle_{exp}$. For our purposes, we will be satisfied with using only one observable. Therefore the MEM distribution becomes

$$P(\mathbf{x}) = \frac{1}{Z} e^{-\lambda f(\mathbf{x})} \sim \frac{1}{Z} e^{-\beta H} \tag{4.2}$$

taking the form of a Boltzmann distribution with λ as an inverse temperature term, and the observable $f(\mathbf{x})$ as the Hamiltonian.

Now let us motivate this paper's MEM describing these 2D flocks. The MEM begins with the assumption that particle interactions are localized (i.e., particles align according to nearby neighbors) and pairwise (e.g. the Hamiltonian has no higher order cross terms, like $\theta_i \theta_j \theta_k$). A first guess for the contribution of one bird to the flock Hamiltonian might look like

$$E_i = - \sum_{j=1}^{N} J_{ij} \cos(\theta_j - \theta_i) \tag{4.3}$$

with J_{ij} nonnegative, and θ representing the direction of bird velocities. Summing the energy contribution of each bird suggests a Hamiltonian,

$$H_1 = \frac{1}{2}\sum_i E_i = -\frac{1}{2}\sum_i^N \sum_j^N J_{ij}\cos(\theta_j - \theta_i), \tag{4.4}$$

where the factor of $\frac{1}{2}$ is included to make a mutual interaction between particles i and j have a total energy contribution of J_{ij}.

Going forward we again adapt the notation that $j \in n^i$ means that j is in the neighborhood set of i. Instead of computing each individual weight J_{ij}, we can further simplify the problem to $J_{ij} = 1$ whenever $j \in n^i$, and 0 otherwise. Another simplification is to assume the size of the neighborhood sets is the same for each particle, so $|n^i| = |n^j| = n$, or equivalently $\sum_j J_{ij} = \sum_k J_{ik}$.

Let us briefly investigate the physical significance of H_1. Note its relation to a type of local correlation:

$$\Phi(\vec{\theta}, J_{ij}) \equiv -\frac{2}{Nn}H_1 = \frac{1}{N}\sum_{i=1}^N \frac{1}{|n^i|}\sum_{j=1}^N J_{ij}\cos(\theta_i - \theta_j) = \frac{1}{N}\sum_{i=1}^N \frac{1}{n}\sum_{j\in n^i} \vec{s}_i \cdot \vec{s}_j. \tag{4.5}$$

In particular Φ measures the average pairwise correlation between each particle i and its neighbors, then averaged over each particle i.

One task remains before we can use the MEM constructed from H_1: to explicitly pick the neighborhood sets n^i for each i. Common schemes to acquire neighborhood sets include topological nearest neighbor and metric interactions [11]. We will find that with another few assumptions, we can conveniently compute many quantities analytically with the nearest neighbor interactions, so we will proceed this way for the sake of intuition.

In full then, we have $J_{ij} = 1$ whenever j is one of the closest n neighbors to i ($J_{ii} = 0$), using the usual metric on 2D periodic boundary conditions, and 0 otherwise. We will find that in general the n neighborhoods are asymmetric, so we switch to a symmetrized Hamiltonian to make our analytic work a bit easier. The Hamiltonian is

$$\Psi(\vec{\theta}) = \frac{1}{2}\sum_{i,j=1}^{N} n_{ij}\cos(\theta_i - \theta_j) = \frac{Nn}{2}\Phi(\vec{\theta}) \tag{4.6}$$

where $n_{ij} \equiv \frac{1}{2}(J_{ij} + J_{ji})$. This means $n_{ij} = 1$ whenever $i \in n^j$ and $j \in n^i$ (mutual neighbors), $n_{ij} = 1/2$ if one is in the neighborhood of the other but not vice versa, and 0 otherwise.

Plugging in the observable Ψ to the maximum entropy distribution and renaming $J \equiv -\lambda$, we have the distribution of interest

$$P_{MEM}(\vec{\theta}, J) = \frac{1}{Z}e^{J\Psi(\vec{\theta},n)}. \tag{4.7}$$

We have chosen $J = -\lambda$, reminiscent of interaction strength of a Boltzmann distribution with $\beta = 1$.

By choosing $\Psi(\vec{\theta}, n)$ as the Hamiltonian, the directions of motion $\{\theta_i\}$ are the only degrees of freedom in the MEM. That is to say, the MEM distribution is conditional on the form of n_{ij}, which is computed from the positional information in the flock snapshot $\{\vec{r_i}\}$ and n. There are two remaining unknown parameters: J and n. J is computed by analytically computing the partition function and using $\langle\Psi\rangle_{MEM} = \partial_J \ln Z = \langle\Psi\rangle_{exp}$ to solve for J. As for the neighborhood size, n, J is computed for all possible n, and are plugged into $\langle\log P_{MEM}\rangle_{exp}$. The value of n that maximizes this likelihood is chosen.

The next task is to compute the partition function. We transform to coordinates θ_i measured with respect to the average direction of the flock, $\theta_{av} \equiv \text{Arg}[\sum_i^N \vec{s}_i]$. This is enforced in the partition function integral by including a delta function term:

$$Z = \int d\vec{\theta} \exp\left[\frac{J}{2}\sum_{ij} n_{ij}\cos(\theta_i - \theta_j)\right] \delta\left(\sum_i \theta_i\right). \tag{4.8}$$

We employ the following simplification

$$Z \approx \int d\vec{\theta} \exp\left[\frac{J}{2}\sum_{ij} n_{ij}\left(1 - \frac{1}{2}(\theta_i - \theta_j)^2\right)\right] \delta\left(\sum_i \theta_i\right) \tag{4.9}$$

which is the spin-wave approximation [15]. This approximation was originally used in the context of ferromagnets in 3D and is justified in the context of this paper as a modeling choice so long as the flocks studied are magnetized. We can evaluate pieces of the sum

$$\begin{aligned} Z &\approx \int d\vec{\theta} \exp\left[\frac{J}{2}\left(\sum_{ij} n_{ij} + \sum_{ij} \theta_i n_{ij} \theta_j - \sum_{ik} n_{ik}\theta_k^2\right)\right] \delta\left(\sum_i \theta_i\right) \\ &= \int d\vec{\theta} \exp\left[\frac{J}{2} N n_c - \frac{J}{2}\vec{\theta}^T \hat{M} \vec{\theta}\right] \delta\left(\sum_i \theta_i\right) \end{aligned} \tag{4.10}$$

where $\hat{M}$ is a matrix with elements $M_{ij} = \delta_{ij}\sum_k n_{ik} - n_{ij}$, and we used $n_{ij} = \frac{1}{2}(J_{ij} - J_{ji})$ to evaluate $\sum_{ij} n_{ij} = \sum_{ij} J_{ij} = \sum_j |n^i| = Nn$. By construction $\hat{M}$ is both symmetric and diagonally dominant ($m_{ii} \geq \sum_{i\neq j} |m_{ij}|$ for all i). Using the Gershgorin circle theorem, we know every eigenvalue of $\hat{M}$ lies within the union of all Gershgorin discs: closed, circular subsets of $\mathbb{C}$ each centered at m_{ii} with radii $\sum_{j\neq i} |a_{ij}|$ respectively [16]. Recalling that $m_{ii} \geq \sum_{i\neq j} |m_{ij}|$ for all i, we know that no part of any disc lies on the left half of the complex plane. Since $\hat{M}$ is symmetric, all eigenvalues are real, so this translates to the statement that $\hat{M}$ has non-negative eigenvalues.

Since it is symmetric, $\hat{M}$ is also diagonalizable via an orthogonal transformation, $\hat{M} = O^T D O$. This allows the integral to be evaluated over its

eigenmodes $\phi_i = O_{ik}\theta_k$

$$\begin{aligned} Ze^{-JNn_c/2} &\approx \int_{-\pi}^{\pi} \prod_i^N d\theta_i \exp\left[-\frac{J}{2}\vec{\theta}^T O^T DO\vec{\theta}\right] \delta\left(\sum_i \theta_i\right) \\ &= \prod_{k=2}^{N} \int_{\phi_k^-}^{\phi_k^+} d\phi_k \, \exp\left[-\frac{J}{2}\vec{\phi}^T D\vec{\phi}\right] \\ &\approx \prod_{k=2}^{N} \int_{-\infty}^{\infty} d\phi_k \exp\left[-\frac{J}{2}\sum_k \lambda_k \phi_k^2\right] = \left(\sqrt{\frac{2\pi}{J}}\right)^{N-1} \frac{1}{\sqrt{\prod_{k=2}^N \lambda_k}} \\ &= \left(\frac{2\pi}{J}\right)^{\frac{N-1}{2}} \frac{1}{\sqrt{\det M^*}} \end{aligned} \tag{4.11}$$

or in full,

$$Z = e^{JNn_c/2}\left(\frac{2\pi}{J}\right)^{\frac{N-1}{2}} \frac{1}{\sqrt{\det M^*}}. \tag{4.12}$$

Details of the integration are explained in Appendix B. We have assumed here that the nullspace of M is one, that is to say, the Hamiltonian only has one symmetry (the global rotation mode), which is true for sufficiently large n. This is a justified approximation in this context because the existence of a second symmetric mode would mean the existence of another rotation at zero energy cost, which would destroy the ordering of the MEM flock. In terms of a biological model, we would have the unacceptable result that the generative flock snapshot could be broken down into non-interacting sub-flocks that just happen to align at the time of the snapshot. Thus n so small that the nullspace of M is greater than 1 are not considered any further into the

calculation. Practically this usually amounts to requiring $n > 3$ in the flocks simulated in this paper.

Now that we have explicitly computed the partition function, we enforce the experimental constraint

$$\langle \Psi \rangle_{exp} = \partial_J \ln Z = \partial_J \left(JNn_c/2 + \frac{N-1}{2} \ln(2\pi/J) - \frac{1}{2} \ln \det M^* \right) = Nn_c/2 - \frac{N-1}{2J}. \tag{4.13}$$

Substituting $\Psi = \frac{Nn_c}{2}\Phi$, we finally arrive at our equation for J:

$$\frac{1}{J} = \frac{N}{N-1} n_c (1 - \langle \Phi \rangle_{exp}). \tag{4.14}$$

The final step is to plug this formula back into P^* and maximize its log likelihood with respect to n, $\langle \ln P^* \rangle_{exp}$. This quantity is

$$\begin{aligned}
\langle \ln P^* \rangle_{exp} &= \langle \ln \frac{1}{Z} e^{J\Psi} \rangle_{exp} = -\ln Z + J\langle \Psi \rangle_{exp} \\
&= -\left(JNn_c/2 + \frac{N-1}{2} \ln(2\pi/J) - \frac{1}{2} \ln \det M^* \right) + J\langle \Psi \rangle_{exp} \\
&= -\frac{1}{2} \ln \left[\left(\frac{2\pi}{J} \right)^{N-1} \right] + \frac{1}{2} \ln \det M^* + J \left(\langle \Psi \rangle_{exp} - \frac{Nn_c}{2} \right) \\
&= \frac{1}{2} \ln \left(\prod_{k=2}^{N} \frac{J\lambda_k}{2\pi} \right) - \frac{JNn_c}{2}(1 - \langle \Phi \rangle_{exp}) \\
&= \frac{1}{2} \sum_{k=2} \ln \frac{\lambda_k (N-1)}{2\pi N n_c (1 - \langle \Phi \rangle_{exp})} - \frac{N-1}{2} \\
&\sim \sum_{k=2} \ln \frac{\lambda_k}{n_c (1 - \langle \Phi \rangle_{exp})}
\end{aligned} \tag{4.15}$$

This quantity is maximized numerically over the n, the n that maximizes $\mathcal{L}$ taking the name n^*.

Since this paper's MEM is constructed from a single snapshot of a Vicsek flock, it is natural to wonder how different snapshots of the same flock can be synthesized together into a single estimate for the interaction parameters. This paper considers two ways in which this may be accomplished. For J_i^* and n_i^* representing the single-frame optimal values of the ith snapshot, the first idea is just to take the average over all snapshots, $\langle J^* \rangle$ and $\langle n^* \rangle$. The second method is to take the likelihoods of each frame $\mathcal{L}_i(n_g, J_g)$ (with n_g

and J_g fixed over all snapshots), average over them, and then maximize this average with respect to n_g and J_g, named a "global" model [3]. For this paper's MEM, it turns out the estimate J_g is simply the harmonic mean of the J_i^*. Since harmonic means are never larger than arithmetic means, we note that $J_g \leq \langle J^* \rangle$. After J_g is computed, n_g is found by maximizing the log likelihood averaged over the snapshots. A full calculation of J_g and n_g is provided in Appendix D.

5 Simulations and Discussion

For simplicity the Vicsek flocks simulated will take interactions of a form similar to that of the MEM Hamiltonian. Recall the update term in θ for the Vicsek model $\langle \theta_{\{n^i\}} \rangle = \arctan\left(\frac{\sum_j m_{ij}\sin(\theta_j)}{\sum_j m_{ij}\cos(\theta_j)}\right)$. For our purposes we will rewrite the Vicsek update equation as

$$\theta_i(t+1) = \arctan\left(\frac{\sin(\theta_i) + \sum_j n_{ij}\sin(\theta_j)}{\cos(\theta_i) + \sum_j n_{ij}\cos(\theta_j)}\right) + \eta_i(t) \tag{5.1}$$

where n_{ij} is the symmetric topological interaction matrix from the previous section (i.e. taking value 1 if mutual nth neighbors, 1/2 if asymmetric nth neighbors, 0 otherwise) for some number of Vicsek neighbors n_V.

We look to see how well the MEM infers the interaction range through maximizing $\mathcal{L}$ as well as plot various collections of flocks looking for and rationalizing trends.

The code used to generate the Vicsek flocks begins the simulation with its particles in random positions and orientations. Each data point in this section's diagrams is either a global or arithmetic average over 30 snapshots, taken at intervals of 50 time steps. For reference the Vicsek flocks typically displayed equilibrium behavior (e.g., reached average $M \equiv \left|\sum_i^N \vec{s_i}\right|$and $\Psi(n_V)$ values) in fewer than 50 snapshots from a fully randomized initial state.

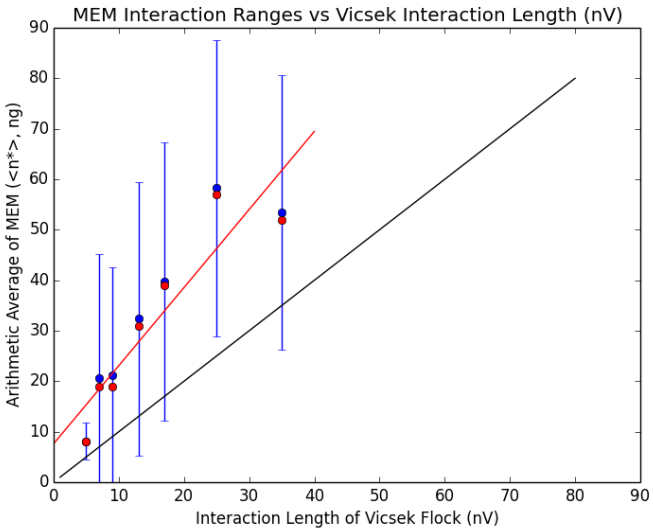

Figure 2: MEM and Vicsek interaction parameters plotted against each other. Each Vicsek flock shares parameters $N = 100$, $\eta = 0.7$, $\nu = 0$. The points in blue are $\langle n_i^* \rangle$ with error bars equal to their respective standard deviations. The points in red correspond to the n_g global interaction estimate. The black line has slope 1, corresponding to what would be a perfect MEM estimate. The red line is an unweighted, linear least squares regression of the n_g and has slope 1.55.

Figure 2 shows two MEM interaction ranges, n_g and $\langle n^* \rangle$ plotted against the actual n_V used in the simulation. It appears n_g and $\langle n^* \rangle$, are quite close to each other, but both significantly overestimate the actual n_V value. For a flock with rapidly changing interaction neighborhoods, we expect the MEM to compensate for its static neighborhoods by overestimating the interaction range [3] [5] [11]. This figure, however, has Vicsek velocity set $\nu = 0$, and since the Vicsek neighborhoods are static, such an argument does not apply. This might be explained intuitively by comparing the alignment terms in the Vicsek dynamics compared to a "plausible" dynamics of the MEM, the Langevin equation with alignment term $\frac{\partial H}{\partial \theta_i} = \sum_j n_{ij} \sin(\theta_j - \theta_i)$ [3]. The

Vicsek alignment term is a vector sum over spins, after which the argument is then taken, whereas the Langevin alignment term is a sum of the perpendicular components to θ_i. Perhaps this difference in alignment terms could be partly responsible for the MEMs tendency to prefer $n_g > n_V$.

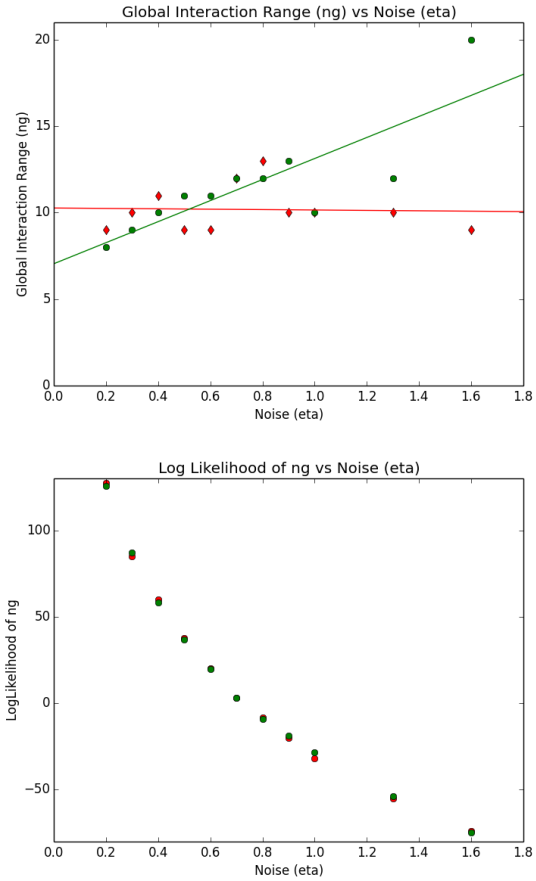

Figure 3: In both figures the red diamonds correspond to an immobile Vicsek flock, and the green circles correspond to a flock with $\nu = 0.05$.(A) n_g plotted against Vicsek noise η. (B) The global log likelihood of the MEM, $\ln(\mathcal{L}(n_g, J_g))$, plotted against Vicsek noise η.

The diagrams in Figure 3 show n_g and its log likelihood plotted against the noise of the Vicsek flock. In Figure 3A both mobile and immobile flocks are plotted on top of each other and are given a linear least squares fit. The data seems noisy, but there is a weak trend suggesting that higher Vicsek noise results in a larger inferred interaction length for the mobile flock. Figure 3B is easier to understand intuitively; it says that higher noise in the Vicsek flock translates to a less accurate reproduction of the n^* correlation. The points are quite smooth, and this is explained by going back to the definition $\log \mathcal{L}_i = \sum_k \frac{\lambda_k}{n(1-\langle\Phi(n)\rangle_{exp})}$. The only term that noise can affect is $\langle\Phi(n)\rangle_{exp}$. As the noise increases the log likelihood decreases as $\frac{1}{1-\langle\Phi(n)\rangle}$ for a fixed n. Since n_g is computed by averaging over each $\mathcal{L}_i$ for a fixed J_g, it's perhaps not so surprising that the graph appears this smooth.

In Figures 4A and 4B there is a weak trend of n_g increasing and J_g de-

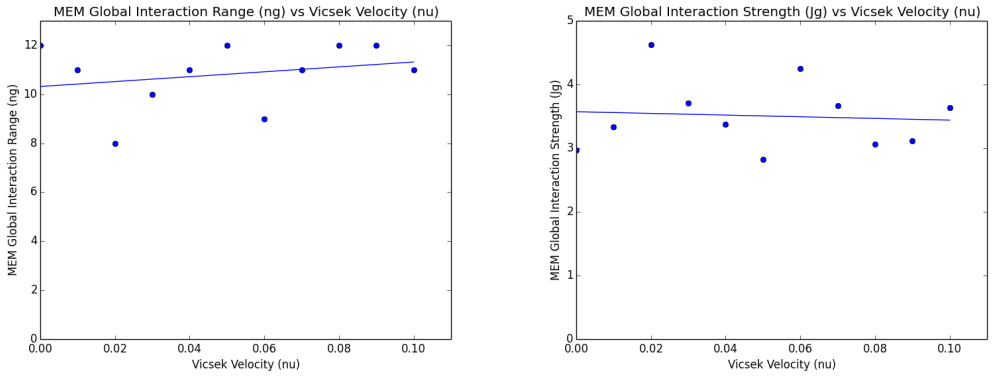

Figure 4: MEM global interaction parameters plotted against noise. The fits for both are linear least square regressions, the slope for n_g against ν is 10.0, the slope for J_g against ν is -1.3.

creasing with flock speed ν. The inferred interaction range increasing with flock velocity is in agreement with the literature, although it is admittedly very noisy [3] [5]. The negative slope of J_g might be in part a byproduct of n_g increasing; we expect n_g increasing means many of the n_i^* increase. According to $\frac{1}{J} \approx n(1 - \langle \Psi(n) \rangle_{exp})$, if there is a small increase in n and if $\langle \Psi(n + \Delta n) \rangle_{exp} \approx \langle \Psi(n) \rangle_{exp}$, then we expect each J_i^* to decrease, and consequently so too J_g.

6 Conclusions

This paper has demonstrated a maximum entropy model designed to describe flocks in 3D enjoy some limited success in 2D. There are many ways to improve these results, but first let us reflect on what has been accomplished. A maximum entropy model, which was chosen specifically for its failure to describe the experimental flock if the inferred interaction length was too low, successfully avoids this problem when the n^* that maximizes $\mathcal{L}(, \mathcal{J})$ is chosen. The n^* estimates, however, are quite noisy in the data collected in this paper. Despite this, the interaction parameters scale more or less as expected.

Opportunities for further research to improve the understanding of the MEM are now proposed. One idea is to increase the number of particles, N, and/or the number of snapshots N_s. The values chosen in the data section of this paper, $N = 100$ and $N_s = 30$, were only chosen because it was the upper limit of what could be comfortably simulated on the hardware available. The choice of $N_s = 30$ seemed inadequately small, since anywhere between 3 and 7 of the 30 snapshots in a given flock had $n^* > 50$. Such a large n^* made J^* significantly smaller. The global interaction strength,

$$\frac{1}{J_g} = \frac{\sum_i^{N_s} \frac{1}{J_i^*}}{N_s} \tag{6.1}$$

is dominated by the minimum of its arguments, and thus was especially sensitive to these outlying low J_i^*. Noisy J_g then resulted in noisy estimates for n_g.

It may also be fruitful to compare averages of observables from experiment to their averages over the MaxEnt distribution. By comparing these averages one could test maximum entropy models by comparing the accuracy of the MaxEnt's averages to experiment.

In conclusion, this thesis has studied an MEM constructed from local interactions applied to a 2D ordered flock. The MEM's interaction parameters were tested against those that generated the experimental flock, and although noisy, the parameter trends reaffirm the intuitions existing in the literature.

7 Appendix A: Details of the Partition Function Integral

After the spin-wave approximation and an orthogonal diagonalization of $\hat{M} = O^T D O$, the partition function is written as

$$Z = e^{JNn_c/2} \int_{-\pi}^{\pi} \prod_{i=1}^{N} d\theta_i \exp\left[-\frac{J}{2}\vec{\theta}^T O^T D O \vec{\theta}\right] \delta\left(\sum_i \theta_i\right). \tag{7.1}$$

Introduce new angular coordinates $\phi_i = O_{ik}\theta_k$. Note that the vector $\frac{1}{\sqrt{N}}(1,\ldots,1)$ is an eigenvalue of $\hat{M}$ with eigenvalue 0. Order the eigenvalues in ascending order, with the $\vec{\theta} = \frac{1}{\sqrt{N}}(1,\ldots,1)$ corresponding to the $\lambda_1 = 0$ eigenvalue. The Jacobian matrix $\frac{\partial \phi_i}{\partial \theta_j}$ is simply equal to O_{ij}, and its absolute value determinant is 1 since O is orthogonal. The integral transforms as

$$Z = e^{JNn_c/2} \int_{\phi_1^-}^{\phi_1^+} \cdots \int_{\phi_N^-}^{\phi_N^+} \prod_i^N d\phi_i \exp\left[-\frac{J}{2}\vec{\phi}^T D \vec{\phi}\right] = e^{JNn_c/2} \prod_{i=1}^{N} \int_{\phi_i^-}^{\phi_i^+} d\phi_i \exp\left[-\frac{J}{2}\lambda_i \phi_i^2\right]. \tag{7.2}$$

Since the bound of integration in θ_i coordinates is simply the N-cube $[-\pi,\pi]\times\cdots\times[-\pi,\pi]$, and a transformation by orthogonal O is equivalent to a sequence of rotations and reflections, the boundary of integration over ϕ_i is another N-cube, but not in general aligned with the ϕ_i coordinate axes. Since the λ_i are nonnegative, we wish to integrate the positive λ_i as decoupled Gaussian integrals.

The approximation made in taking these integrals to be Gaussian is that the boundaries of integration $|\phi_i^{\pm}|$ are much larger than the standard deviation of the Gaussians, $\frac{1}{\sqrt{J\lambda_i}}$. Note that $|\phi_i^{\pm}| \geq \pi$, since before the rotation by O, the closest distance from the origin to the N-cube boundary is π. The integral with the largest non-zero variance is that with the smallest non-zero $\lambda = \lambda_{min}$, so the simulation computes $\pi\sqrt{J\lambda_{min}}$, the minimum possible length to the boundary in units of standard deviation, σ. In the flocks tested $\pi\sqrt{J\lambda_{min}}$ was always > 2, and usually > 3, so the approximation appears to be fairly accurate.

8 Appendix B: Global Model

Again following the footsteps of Bialek and collaborators' 2012 paper, we recognize that the MEMs under consideration only use data from single frames, but ultimately describe a changing flock. We may reconcile this to some degree by introducing the additional assumption that the interaction parameters (J and n_c) are the same throughout every snapshot of the same flock, add the likelihoods of each frame, and pick the "global" interaction strength that maximizes this sum log likelihood. Once the J is computed, we repeat this process to find the global interaction length, n_g.

The likelihood the MEM produces the snapshot at time t was computed earlier ($\mathcal{L}_t$), and we sum over all N_s number of snapshots for the total likelihood. We hold n fixed through this first maximization.

$$\mathcal{L}_{tot} = \sum_{t=1}^{N_s} \mathcal{L}_t = \sum_t \langle \ln(P_{MEM}) \rangle_t = \sum_t \left[J \langle \Psi(n_c, t) \rangle_{exp} - \ln Z(J, n_c, t) \right] \quad (8.1)$$

The extrema with respect to J satisfy

$$\left. \frac{\partial \mathcal{L}_{tot}}{\partial J} \right|_{J=J_{gl}} = 0 = \sum_t \left(\langle \Psi(n_c, t) \rangle_{exp} - \partial_J \ln Z(J, t) \Big|_{J=J_{gl}} \right). \quad (8.2)$$

From the second section of this thesis, we showed the equation

$$\langle \Psi(n_c) \rangle_{exp} = \partial_J \ln Z(J, t, n_c) \quad (8.3)$$

is satisfied whenever we pick $J = J^*$, the J that maximizes the log likelihood of reproduces that snapshot. So rearranging (10.2), we have

$$\sum_t \partial_J \ln Z(J, t, n_c)\Big|_{J=J_{gl}} = \sum_t \partial_J \ln Z(J, t, n_c)\Big|_{J=J_t^*}. \tag{8.4}$$

The last step is to plug in the evaluated derivative of the log partition function: $\partial_J \ln Z = \frac{Nn_c}{2} - \frac{N-1}{2J}$, so we cancel all but the J terms to find

$$\sum_t \frac{1}{J_{gl}} = \sum_t \frac{1}{J_t^*} \tag{8.5}$$

giving final answer

$$\frac{1}{J_{gl}} = \frac{1}{N_s} \sum_{t=1}^{N_s} \frac{1}{J_t^*}. \tag{8.6}$$

In conclusion our "global" interaction strength J_{gl} turns out to be simply the harmonic mean of the J's that optimize the MEM for each individual frame.

Since J_{gl} is not a function of n_c^{gl}, it is a simple computational matter to compute $\mathcal{L}(J_{gl}, n_c)$ for each possible value of n_c, finding the maximum likelihood by brute force.

References

[1] Couzin ID, Kraus J (2003) Self-organization and collective behavior in vertebrates. *Adv Study Behav* 32: 1-75

[2] Vicsek T, Czirok A, Ben-Jacob E, Cohen I, Shochet O (1995) Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys Rev Lett* 75: (6) 1226-1229

[3] Bialek W, et al. (2012) Statistical mechanics for natural flocks of birds. *PNAS* 109: (13) 4786-4791

[4] Bialek W, et al. (2014) Social interactions dominate speed control in poising natural flocks near criticality. *PNAS* 111: (20) 7212-7217

[5] Toner J, Tu Y (1995) Long-Range Order in a Two-Dimensional Dynamical XY Model: How Birds Fly Together. *Phys Rev Lett* 75: (23) 4326-4329

[6] Mora, T. & Bialek, W. J Stat Phys (2011) 144: 268. https://doi.org/10.1007/s10955-011-0229-4

[7] Shannon CE (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal* Vol. 27, pp. 379-423, 623-656

[8] Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620-630

[9] Mora T, Walczak AM, Bialek W, Callan CG (2010) Maximum entropy models for anti-body diversity. *Proc Natl Acad Sci USA* 107:5405-5410

[10] Meshulam L, Gauthier JL, Brody CD, Tank DW, Bialek W (2016) Collective behavior of place and non-place neurons in the hippocampal network. arXiv:1612.08935v1 [q-bio.NC]

[11] Cavagna A, Giardina I, Ginelli F, Mora T, Piovani D, Tavarone R, Walczak AM (2014) Dynamical maximum entropy approach to flocking. *Phys Rev* 89: (4) 042707

[12] Kittel C *Introduction to Solid State Physics* (John Wiley & Sons Inc., Hoboken, NJ)

[13] Mermin ND, Wagner H (1966) Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models. *Phys Rev Lett* 17: (22) 1133-1136

[14] Cover TM, Thomas JA *Elements of Information Theory* (John Wiley & Sons Inc., Hoboken, NJ)

[15] Dyson FJ (1956) General theory of spin-wave interactions. Phys Rev 102:1217-1230

[16] Horn RA, Johnson CR *Matrix Analysis* (Cambridge Univ Press, New York, NY)