

Learning in the Absence of Training Data – a Galactic Application

Cedric Spire, Dalia Chakrabarty

Abstract There are multiple real-world problems in which training data is unavailable, and still, the ambition is to learn values of the system parameters, at which test data on an observable is realised, subsequent to the learning of the functional relationship between these variables. We present a novel Bayesian method to deal with such a problem, in which we learn a system function of a stationary dynamical system, for which only test data on a vector-valued observable is available, and training data is unavailable. This exercise borrows heavily from the state space probability density function (*pdf*), that we also learn. As there is no training data available for either sought function, we cannot learn its correlation structure, and instead, perform inference (using Metropolis-within-Gibbs), on the discretised form of the sought system function and of the *pdf*, where this *pdf* is constructed such that the unknown system parameters are embedded within its support. Likelihood of the unknowns given the available data, is defined in terms of such a *pdf*. We make an application to the learning of the density of all gravitational matter in a real galaxy.

1 Introduction

The study of rich correlation structures of high-dimensional random objects, is often invoked when learning the unknown functional relationship between an observed random variable, and some other parameters that might inform on the properties of a system. A problem in which a vector of system parameters (say $\boldsymbol{\rho} \in \mathcal{R} \subseteq \mathbb{R}^p$) is related to an observed response variable (say $\mathbf{Y} \in \mathcal{Y} \subseteq \mathbb{R}^d$), is easily visualised by the

Cedric Spire

Loughborough University. Department of Mathematical Sciences. e-mail: c.spire@lboro.ac.uk

Dalia Chakrabarty

Loughborough University. Department of Mathematical Sciences. e-mail: d.chakrabarty@lboro.ac.uk

arXiv:1811.09204v1 [stat.AP] 22 Nov 2018

equation: $\mathbf{Y} = \boldsymbol{\xi}(\boldsymbol{\rho})$, where $\boldsymbol{\xi} : \mathcal{R} \rightarrow \mathcal{Y}$. Given training data $\mathbf{D} = \{(\boldsymbol{\rho}_i, \mathbf{y}_i)\}_{i=1}^{N_{data}}$, we aim to learn this unknown mapping $\boldsymbol{\xi}(\cdot)$ within the paradigm of *supervised learning*. By "training data" we mean here the pairs composed of chosen design points $\boldsymbol{\rho}_i$, and the output \mathbf{y}_i that is generated at $\boldsymbol{\rho}_i$; $i = 1, \dots, N_{data}$. Methods to perform supervised learning are extensively covered in the literature [12, 16, 17, 18]. Having learnt $\boldsymbol{\xi}(\cdot)$, one could use this model to predict the value $\boldsymbol{\rho}$ [5], at which the test datum \mathbf{y}_{test} on \mathbf{Y} is realised – either in the conventional framework as $\boldsymbol{\rho} = \boldsymbol{\xi}^{-1}(\mathbf{Y})|_{\mathbf{Y}=\mathbf{y}_{test}}$, or as the Bayesian equivalent. Such prediction is possible, only subsequent to the learning of the functional relation between $\boldsymbol{\rho}$ and \mathbf{Y} using training data \mathbf{D} .

However, there exist physical systems for which only measurements on the observable \mathbf{Y} are known, i.e. training data is not available. The disciplines affected by the absence of training data are diverse. In engineering [20], anomaly detection is entirely sample-specific. There is no training data that allows for the learning of a functional relationship between anomaly occurrence (parametrised by type and severity of anomaly) and conditions that the sample is subjected to. Yet, we need to predict those anomalies. In finance, such anomalies in stock price trends are again outside the domain of supervised learning, given that the relationship between the market conditions and prices have not been reliably captured by any "models" yet. In neuroscience, [1], a series of neurons spike at different amplitudes, and for different time widths, to cause a response (to a stimulus). We can measure the response's strength and the parameters of firing neurons, but do not know the relation between these variables. Again, in petrophysics, the system property that is the proportion of the different components of a rock (eg. water, hydrocarbons), affects Nuclear Magnetic Resonance (NMR) measurements from the rock [7, 21]. However, this compositional signature cannot be reliably estimated given such data, using available estimation techniques. Quantification of petrological composition using the destructive testing of a rock, is highly exclusive, and expensive, to allow for a sample that is large enough to form a meaningful training data set. Also, the resulting training data will in general be unrepresentative of any new rock, since the relationship between the (compositional) system property and (NMR) observable is highly rock-specific, being driven by geological influences on the well that the given rock is obtained from. Therefore any training data will need to be substantially diverse, and as stated before, this is unachievable in general. Equally, this dependence on the latent geological influence annuls the possibility of using numerical simulations to generate NMR data, given design compositional information. Thus, generation of training data is disallowed in general.

In this work, we advance the learning of the sought functional relation between an observable and a system parameter vector, in such a challenging (absent training) data situation; this could in principle, then be undertaken as an exercise in *unsupervised learning*, though opting for the more robust supervised learning route is still possible, as long as the missing training data is generated, i.e. we are able to generate the $\boldsymbol{\rho}_i$ at which the measured (test) datum, \mathbf{y}_i on \mathbf{Y} , is available, $\forall i \in \{1, \dots, N_{data}\}$. Our new method for accomplishing this, is to invoke a system property that helps link $\boldsymbol{\rho}$ with \mathbf{Y} , and this is possible in physical systems for which we have – at least partial – observed information. To clarify, what we advance in the face of the absent

training data, is the pursuit of the probability density function of the observable \mathbf{Y} , on which data is available, and employ this to learn the system parameter vector $\boldsymbol{\rho}$. We undertake such an exercise in a Bayesian framework, in which we seek posterior of the *pdf* of the observables, and the system parameters, given the available data.

The sought parameter vector could inform on the behaviour, or structure, of the system (eg. it could be the vectorised version of the density function of all gravitating matter in a distant galaxy). The state space *pdf* establishes the link between this unknown vector, and measurements available on the observable (that may comprise complete or incomplete information on the state space variable). We consider dynamical systems, s.t. the system at hand is governed by a kinetic equation [11]; we treat the unknown system parameter vector as the stationary parameter in the model of this dynamical system. In the novel Bayesian learning method that we introduce, this parameter is embedded within the support of the state space *pdf*. We describe the general model in Section 2, that is subsequently applied to an astronomical application that is discussed in Section 3. Inference is discussed in Section 4, where inference is made on the state space *pdf* and the sought system parameters, given the data that comprises measurements of the observable, using Metropolis-within-Gibbs. Results are presented in Section 5, and the paper is rounded up with a conclusive section (Section 6).

2 General Methodology

We model the system as a dynamical one, and define the state space variable as a p -dimensional vector $\mathbf{S} \in \mathcal{S} \subseteq \mathbb{R}^p$. Let the observable be $\mathbf{Y} \in \mathcal{Y} \subseteq \mathbb{R}^d$; $d < p$, such that only some (d) of the p different components of the state space vector \mathbf{S} can be observed. In light of this situation that is marked by incomplete information, we need to review our earlier declaration of interest in the probability density function of the full state space vector. Indeed, we aim to learn the *pdf* of the state space variable \mathbf{S} , and yet, have measured information on only \mathbf{Y} , i.e. on only d of the p components of \mathbf{S} . Our data is then one set of measurements of the observable \mathbf{Y} , and can be expressed by $\mathbf{D} = \{\mathbf{y}^{(k)}\}_{k=1}^{N_{data}}$. If the density of \mathcal{S} is to be learnt given data on \mathbf{Y} , such incompleteness in measured information will have to be compensated for by invoking some independent information. Such independent information is on the symmetry of \mathcal{S} .

It follows that unobserved components of \mathbf{S} will have to be integrated out of the state space *pdf*, in order to compare against data that comprises measurements of the observables. This state space *pdf* that the unobserved variables are integrated out of, is equivalently projected onto the space \mathcal{Y} of observables, and therefore, we refer to it as the *projected state space pdf*. The likelihood of the model parameters, given the data, is simply the product of the projected state space *pdf* over all the data points. But until now, the unknown model parameters have not yet appeared in our expression of the likelihood. The next step is then to find a way for embedding the sought system parameters, in the support of the projected state space *pdf*.

This can be achieved by assuming that our dynamical system is stationary, so that its state space *pdf* does not depend on time-dependent variables. In other words, the rate of change of the state space *pdf* is 0. This allows us to express the *pdf* as dependent on the state space vector \mathbf{S} , but only via such functions of (some or all amongst) S_1, \dots, S_p that are not changing with time; in fact, the converse of this statement is also true. This is a standard result, often referred to as Jeans Theorem [3, 14]. The model parameters that we seek, can be recast as related to such identified time-independent functions of all/some state space coordinates of motion. Thus, by expressing the state space *pdf* as a function of appropriate constants of motion, we can embed system parameters into the support of the sought *pdf*.

As stated above, this *pdf* will then need to be projected into the space of observables \mathcal{Y} , and we will convolve such a projected *pdf* with the error density, at every choice of the model parameters. Then assuming the data to be *iid*, the product of such a convolution over the whole data set will finally define our likelihood. Using this likelihood, along with appropriate priors, we then define the posterior probability density of the model parameters and the state space *pdf*, given the data \mathbf{D} . Subsequently we generate posterior samples using Metropolis-within-Gibbs. scheme.

We recall that in absence of training data on a pair of r.v.s, we cannot learn the correlation structure of the functional relationship between these variables. In such situations, instead of the full function, we can only learn the vectorised version of the sought function. In other words, the relevant interval of the domain of the function is discretised into a bin, and the value of the function held a constant over any such bin; we can learn the functional value over any such bin.

3 Astrophysics Application

Our astrophysics application is motivated by the wish to learn the contribution of dark matter, to the density function of all gravitating mass in a distant galaxy. While information on light-emitting matter is available, it is more challenging to model the effects of dark matter since, by definition, one cannot observe such matter (as it does not emit/reflect light of any colour). However, physical phenomena such as: the distortion of the path of light by gravitational matter acting as gravitational lenses; temperature distribution of hot gas that is emanating from a galaxy; motions of stars or other galactic particles that is permitted in spite of the attractive gravitational pull of the surrounding galactic matter, allow us to confirm that non-observable, dark matter is contributing to the overall gravitational mass density of the galaxy. In fact, astrophysical theories suggest that the proportion of dark matter in older galaxies (that are of interest to us here) is the major contributor to the galactic mass, over the minor fraction of luminous galactic matter [13]. We can compute the proportion of this contribution, by subtracting the density of the luminous matter from the overall density. It is then necessary to learn the gravitational mass density of the whole system in order to learn the density of dark matter.

We begin by considering the galaxy at hand to be a stationary dynamical system, i.e. the distribution of the state space variable does not depend on time. Let $\mathbf{S} = (X_1, X_2, X_3, V_1, V_2, V_3)^T \in \mathcal{S} \subseteq \mathbb{R}^6$ define the state space variable of a galactic particle, where $\mathbf{X} = (X_1, X_2, X_3)^T$ is defined as its 3-dimensional location vector and $\mathbf{V} = (V_1, V_2, V_3)^T$ as the 3-dimensional velocity vector of the galactic particle. Our data consists of measurement of the one observable velocity coordinate V_3 , and two observable spatial coordinates, X_1, X_2 , of N_{data} galactic particles (eg. stars). That is, for each galactic particle, we have measurements of $\mathbf{Y} = (X_1, X_2, V_3)^T \in \mathcal{Y} \subseteq \mathbb{R}^3$. For N_{data} observations, our data is thus $\mathbf{D} = \{\mathbf{y}^{(k)}\}_{k=1}^{N_{data}}$.

The system function that we are interested in learning here, is the density function $\rho(X_1, X_2, X_3)$ of the gravitational mass of all matter in the considered galaxy, where we assume that this gravitational mass density $\rho(\cdot)$ is a function of the spatial coordinates \mathbf{X} only. This system function does indeed inform on the structure of the galactic system – for it tells us about the distribution of matter in the galaxy; it also dictates the behaviour of particles inside the galaxy, since the gravitational mass density is deterministically known as a function of the gravitational potential $\Phi(X_1, X_2, X_3)$ via the Poisson equation ($\nabla^2 \Phi(X_1, X_2, X_3) = -4\pi G\rho(X_1, X_2, X_3)$, where G is the known Universal Gravitational constant, and ∇^2 is the Laplacian operator), which is one of the fundamental equations of Physics [10]. The potential of a system dictates system dynamics, along with the state space distribution.

Here, we assume that the state space density of this dynamical system does not vary with time, i.e. $\frac{df[X_1(t), X_2(t), X_3(t), V_1(t), V_2(t), V_3(t)]}{dt} = 0$. This follows from the consideration that within a typical galaxy, collisions between galactic particles are extremely rare [3]. We thus make the assumption of a collisionless system evolving in time according to the *Collisionless Boltzmann Equation* (CBE) [3, 6]. As motivated above, this allows us to express the state space *pdf* as dependent on those functions of $X_1, X_2, X_3, V_1, V_2, V_3$ that remain invariant with time, along any trajectory in the state space \mathcal{S} ; such time-invariant constants of motion include energy, momentum, etc. It is a standard result that the constant of motion that the state space *pdf* has to depend on, is the energy $E(X_1, X_2, X_3, \|\mathbf{V}\|)$ of a galactic particle [2, 8], where $\|\cdot\|$ represents the Euclidean norm of a vector. Here, energy is given partly by kinetic energy that is proportional to $\|\mathbf{V}\|^2$, and partly by potential energy, which by our assumption, is independent of velocities. Secondly, given that the state space is 6-dimensional, the number of constants of motion ≤ 5 , in order to let the galactic particle enjoy at least 1 degree of freedom, i.e. not be fixed in state space [8].

We ease our analysis by assuming that the state space *pdf* is a function of energy only. This can be rendered equivalent to designating the symmetry of isotropy to the state space \mathcal{S} , where isotropy implies invariance to rotations in this space, i.e. the state space *pdf* is assumed to be such a function of \mathbf{X} and \mathbf{V} , that all orthogonal transformations of \mathbf{X} and \mathbf{V} preserve the state space *pdf*. The simple way to achieve the equivalence between a isotropic state space *pdf* and the lone dependence on energy E of the *pdf*, is to ensure that the gravitation mass density, (and therefore the gravitational potential) at all points at a given Euclidean distance from the galactic centre, be the same, i.e. the distribution of gravitational mass abides by spherical

symmetry s.t. $\rho(\cdot)$ (and therefore $\Phi(\cdot)$) depends on X_1, X_2, X_3 via the Euclidean norm $\|\mathbf{X}\|$ of the location vector \mathbf{X} of a particle. Then energy E is given as the sum of the $\|\mathbf{V}\|^2$ -dependent kinetic energy, and the $\|\mathbf{X}\|$ -dependent potential energy. Spherical mass distribution is not a bad assumption in the central parts of “elliptical” galaxies that are of interest for us, as these have a global triaxial geometry.

To summarise, state space *pdf* is written as $f(E)$, and we embed $\rho(\cdot)$ into the support of this state space *pdf* $f(E)$, by recalling that energy E is partly the gravitational potential energy $\Phi(\cdot)$ that is deterministically related to the gravitational mass density $\rho(\cdot)$ through Poisson equation.

As there is no training data available to learn the correlation structure of the sought functions $\rho(\mathbf{X})$ and $f(E)$, we can only learn values of these functions at specified points in their domains, i.e. learn their vectorised forms $\boldsymbol{\rho}$ and \mathbf{f} respectively, where $\boldsymbol{\rho} := (\rho_1, \dots, \rho_{N_X})^T$, with $\rho_i = \rho(\mathbf{x})$ for $\mathbf{x} \in [\mathbf{x}_{i-1}, \mathbf{x}_i]; i = 1, \dots, N_X$. The discretised form of $f(E)$ is similarly defined, after discretising the relevant (non-positive) E -values (to indicate that the considered galactic particles are bound to the galaxy by gravitational attraction), into N_E number of E -bins. Then in terms of these vectorised versions of the state space *pdf* likelihood of the unknown parameters $\rho_1, \dots, \rho_{N_X}, f_1, \dots, f_{N_E}$, given data on the observable \mathbf{Y} is:

$$\ell(\boldsymbol{\rho}, \mathbf{f} | \{\mathbf{y}^{(k)}\}_{k=1}^{N_{data}}) = \prod_{k=1}^{N_{data}} v(\mathbf{y}^{(k)}, \boldsymbol{\rho}, \mathbf{f}), \quad (1)$$

where $v(\cdot)$ is the projected state space *pdf*.

We also require that $\rho_1 \geq 0, \dots, \rho_{N_X} \geq 0, f_1 \geq 0, \dots, f_{N_E} \geq 0$, and that $\rho_i \geq \rho_{i+1}, i = 1, \dots, N_X - 1$. The latter constraint is motivated by how the mass in a gravitating system (such as a galaxy) is distributed; given that gravity is an attractive force, the stronger pull on matter closer to the centre of the galaxy, implies that gravitational mass density should not increase, as we move away from the centre of the system. These constraints are imposed via the inference that we employ.

4 Inference

Inference on the unknown parameters – that are the components of $\boldsymbol{\rho}$ and \mathbf{f} – is undertaken using Metropolis-within-Gibbs. In the first block update during any iteration, the $\rho_1, \dots, \rho_{N_X}$ parameters are updated, and subsequently, the f_1, \dots, f_{N_E} parameters are updated in the 2nd block, at the updated ρ -parameters, given the data \mathbf{D} that comprises N_{data} measurements of the observed state space variables X_1, X_2, V_3 that are the components of the observable vector \mathbf{Y} .

Imposition of the monotonicity constraint on the ρ parameters, s.t. $\rho_i \geq \rho_{i+1}, i = 1, \dots, N_X - 1$, renders the inference interesting. We propose ρ_i from a Truncated Normal proposal density that is left truncated at $\rho_{i+1}, \forall i = 1, \dots, N_X - 1$, and propose ρ_{N_X} from a Truncated Normal that is left truncated at 0. The mean of the proposal density is the current value of the parameter and the variance is experi-

mentally chosen, as distinct for each $i \in \{1, \dots, N_X\}$. Such a proposal density helps to maintain the non-increasing nature of the ρ_i -parameters, with increasing i . At the same time, non-negativity of these parameters is also maintained. We choose arbitrary seeds for $\rho_1, \dots, \rho_{N_X}$, and using these as the means, a Gaussian prior is imposed on each parameter. The variance of the prior densities is kept quite large, and demonstration of lack of sensitivity to the prior choices, as well as the seeds, is undertaken. As for components of the vectorised state space *pdf*, there is no corre-

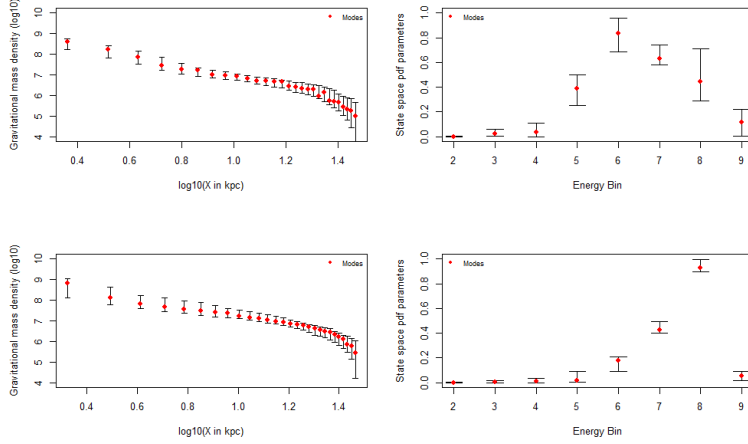


Fig. 1 Results from a MCMC scheme showing the 95% HPDs for all the parameters to learn, for both PNe (top row) and GC (bottom row) data. Modes are shown as red dots. **Top Row:** HPDs on the ρ (left), and the f parameters for the PNe data. **Bottom Row:** HPDs on the ρ (left), and the f parameters for the GC data.

lation information to be enjoyed in this case, unlike in the case of the components of the vectorised gravitational mass density function. We propose f_j from a Truncated Normal (to maintain non-negativity), where the mean of this proposal density is the current value of the parameter and the variance is chosen by hand. Loose Gaussian priors are imposed, while the same seed value is used $\forall j \in \{1, \dots, N_E\}$.

An important consideration in our work is the choice of N_X and N_E . We could have treated these as unknowns and attempted learning these from the data; however, that would imply that the number of unknowns is varying from one iteration to another, and we desired to avoid such a complication, especially since the data strongly suggests values of N_X and N_E . We choose N_X by binning the range of $R_p := \sqrt{X_1^2 + X_2^2}$ values in the data \mathbf{D} , s.t. each resulting R_p -bin includes at least one observed value of V_3 in it, and at the same time, the number of R_p -bins is maximised. Again, we use the available data \mathbf{D} to compute the empirical values of energy E , where an arbitrarily scaled histogram of the observed R_p is used to mimic the vectorised gravitational mass density function, that is then employed to compute the empirical estimate of the vectorised gravitational potential function, that con-

tributes to E values. We admit maximal E -bins over the range of the empirically computed values of E , s.t. each such E -bin contains at least one datum in \mathbf{D} .

5 Results

We have input data on location and velocities of 2 kinds of galactic particles (called ‘‘Globular Clusters’’, and ‘‘Planetary Nebulae’’ – respectively abbreviated as GC and PNe), available for the real galaxy NGC4494. The GC data comprises 114 measurements of $\mathbf{Y} = (X_1, X_2, V_3)^T$, for the GCs in NGC4494 [9]. Our second data set (PNe data), comprises 255 measurements of the PNe [15]. Results of the learnt 95% HPDs for all parameters, given both PNe (top row) and GC (bottom row) data, are shown in Figure 1. Significant inconsistencies between the learnt gravitational mass density parameters can suggest interesting dynamics, such as splitting of the galactic state space into multiple, non-communicating sub-spaces [4], but for this galaxy, it is noted that such parameters learnt from the 2 datasets, concur within learnt HPDs.

6 Conclusions

An astronomical implication of our work is that ρ_1 learnt from either dataset suggests a very high gravitational mass density in the innermost R_p -bin ($\approx 1.6\text{kpc}$), implying gravitational mass $\gtrsim 10^9$ times mass of the Sun, enclosed within this innermost radial bin. This result alone does not contradict the suggestion that NGC4494 harbours a central supermassive blackhole (SMBH) of mass $\sim 2.69 \pm 2.04 \times 10^7$ solar masses [19]. Very interestingly, our results indicate that for both GCs and PNe, most particles lie in the intermediate range of energy values; this is also borne by the shape of the histogram of the empirically computed energy using either dataset, where this empirical E value computation is discussed in the last paragraph of Section 4. However, owing to its intense radially inward gravitational attraction, a central SMBH is expected to render the potential energy (and therefore the total energy E) of the particles closer to the galactic centre, to be much higher negative values, than those further away, while also rendering the number (density) of particles to be sharply monotonically decreasing with radius away from the centre. This is expected to render the energy distribution to be monotonically decreasing as we move towards more positive E values – in contradiction to our noted non-monotonic trend. So while our results are not in contradiction to the report of a very large value of mass enclosed within the inner parts of NGC4494, interpretation of that mass as a SMBH does not follow from our learning of the state space *pdf*.

The learning of the gravitational mass density function, and state space *pdf* – as well as that of the relation $\xi(\cdot)$ between the observable state space coordinates, and the system function/vector – can be done after generating the training dataset relevant to the functional learning problem at hand. Applications in Petrophysics and Finance are also planned.

References

- [1] Ahmad, S., Lavin, A., Purdy, S., and Agha, Z. (2017). *Unsupervised real-time anomaly detection for streaming data*. Number Volume 262 in Neurocomputing.
- [2] Binney, J. (1982). Dynamics of elliptical galaxies and other spheroidal components. *Annual Review of Astronomy and Astrophysics*, 20:399–429.
- [3] Binney, J. and Tremaine, S. (1987). *Galactic Dynamics*. Princeton University Press, Princeton.
- [4] Chakrabarty, D. (2017). A new bayesian test to test for the intractability-countering hypothesis. *Jl. of American Statistical Association*, 112:561–577.
- [5] Chakrabarty, D., Biswas, M., and Bhattacharya, S. (2015). Bayesian nonparametric estimation of milky way parameters using matrix-variate data, in a new gaussian process based method. *Electronic Journal of Statistics*, 9(1):1378–1403.
- [6] Choudhuri, A. (2010). *Astrophysics for Physicists*. Cambridge University Press.
- [7] Coates, G. R., Xhao, L., and Prammer, M. G. (1999). *NMR logging; principles & applications*. Halliburton Energy Services Publication H02308, Houston.
- [8] Contopoulos, G. (1963). A classification of the integrals of motion. *Astrophysical Jl*, 138:1297–1305.
- [9] Foster, C. et al. (2011). Global properties of ordinary early-type galaxies: photometry and spectroscopy of stars and globular clusters in ngc 4494. *Monthly Notices of Royal Astronomical Society*, 415:3393–3416.
- [10] Goldstein, H., Poole, C. P., and Safko, J. (2002). *Classical Mechanics*. Addison-Wesley Longman, Incorporated.
- [11] Gressman, P. T. and Strain, R. M. (2010). Global classical solutions of the boltzmann equation with long-range interactions. *Proceedings of the National Academy of Sciences*.
- [12] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer-Verlag New York.
- [13] Kalinova, V. (2014). *Mass Distributions of Galaxies from SAURON and CALIFA Stellar Kinematic Maps*. Doctoral thesis, Max-Planck-Institut fur Astronomie.
- [14] Merritt, D. (2013). *Dynamics and Evolution of Galactic Nuclei*. Princeton University Press, Princeton.
- [15] Napolitano, N. et al. (2009). The Planetary Nebula Spectrograph elliptical galaxy survey: the dark matter in NGC 4494. *Monthly Notices of the Royal Astronomical Society*, 393:329–353.
- [16] Neal, R. M. (1998). Regression and classification using gaussian process priors (with discussion). In et. al, J. M. B., editor, *Bayesian Statistics 6*, pages 475–501. Oxford University Press.
- [17] Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. The MIT Press, MIT.
- [18] Russel, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Pearson, 3 edition.
- [19] Sadoun, R. and Colin, J. (2012). $M_{BH} - \sigma$ relation between supermassive black holes and the velocity dispersion of globular cluster systems. *Monthly Notices of the Royal Astronomical Society: Letters*, 426(1):L51–L55.
- [20] Sun, X., Yao, H., Ji, R., Liu, X., and Xu, P. (2011). Unsupervised fast anomaly detection in crowds. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 1469–1472, New York, NY, USA. ACM.
- [21] Wang, P., Jain, V., and Venkataramanan, L. (2016). Sparse bayesian t1-t2 inversion from borehole nmr measurements. In *Proceedings of SPWLA 57 Annual Logging Symposium, 25-29 June 2016, Reykjavik*.