Space Object Shape Inversion via Adaptive Hamiltonian Markov Chain Monte Carlo

Richard Linares[*]

University of Minnesota, Twin City Campus, Minneapolis, NM, 55403

John L. Crassidis[†]

University at Buffalo, State University of New York, Amherst, NY, 14260-4400

This paper presents a new approach to estimate an observed space object's shape, while also inferring other attributes such as its inertial attitude and surface parameters. An Adaptive Hamiltonian Markov Chain Monte Carlo estimation approach is employed, which uses light curve data and process inversion to estimate the shape and other attributes. The main advantage of this approach over previous ones is that it can estimate these attributes simultaneously, whereas previous approaches typically rely on *a priori* knowledge of one or more of them to estimate a particular attribute. Also, unlike previous approaches, the new approach is shown to work well for relatively high dimensions and non-Gaussian distributions of the light curve inversion problem. Simulation results involving single and multiple faceted objects with both diffuse and spectral light curve reflections are shown. Good results are obtained for all cases.

I.   Introduction

Space Situational Awareness (SSA) has many definitions depending on the goal at hand, but in general it involves collecting and maintaining knowledge of all space objects (SOs) orbiting the Earth and the space environment. This task is becoming more difficult as the number of objects currently tracked by the U.S. increases due to breakup events and

---

[*]Assistant Professor, Department of Aerospace Engineering and Mechanics. Email: rlinares@umn.edu. Member AIAA.

[†]CUBRC Professor in Space Situational Awareness, Department of Mechanical & Aerospace Engineering. Email: johnc@buffalo.edu. Fellow AIAA.

improving tracking capabilities [1]. The Space Surveillance Network (SSN) is tasked with maintaining information on over 22,000 objects, 1,100 of which are active, with a collection of optical and radar sensors. Determining physically significant characteristics, i.e. attributes, that go beyond simple orbital states is a key objective which is required for protecting space capabilities and achieving SSA. For example, the SSN catalog currently includes radar cross-section and a non-conservative force parameter, analogous to a ballistic coefficient, which provide additional SO characterization information beyond position and velocity. Future SSA systems will have to be capable of building a much more detailed picture of SO attributes in order to maintain better knowledge of their characteristics, which ultimately may lead to better tracking capabilities. For example, current SOs are modeled as "cannonballs" with uniform surface material characteristics. Thus, orbit/attitude effects are ignored with these models. Knowing the shape, attitude and surface parameters can lead to better propagation using detailed orbit/attitude models, which ultimately leads to better tracking capabilities and SSA.

Traditionally, optical tracking data are primarily used for orbit determination of most SOs. These measurements have also been shown to provide information on SO attributes because they depend on shape [2,3], attitude [4], angular velocity [2,5–7], and surface parameters [8]. These quantities are determined through processing light curve measurements, which are the time history of the observed brightness derived from non-resolved images of SOs. Although resolved imaging of some classes of SOs is possible with powerful ground-based telescopes, such as the Air Force Maui Optical and Supercomputing site Advanced Electro-Optical System, a large class of objects are too small and/or too distant to be fully resolved. For example, most operational SOs in geosynchronous orbits as well as "micro" and "nano" satellites in any orbit are too small and/or too distant to be resolved. Extracting SO attributes from non-resolved data involves solving a nonlinear, weakly observable, and (oftentimes) non-Gaussian inverse problem. However, solving this inverse problem for SO attributes from light curve data can provide improved information for a very critical class of SOs.

Light curve inversion has a rich history in the asteroid characterization community, with a number of successful software solutions [9,10]. The inversion approaches used for asteroids are not easily transferrable to satellite applications because these approaches make a number of assumptions that are not valid for satellites. For example, the asteroids are assumed to reflect light in a purely Lambertian manner, which is not a valid assumption for artificial SOs which can exhibit considerable specular reflections, i.e. glints. Moreover, the rotational dynamics assumed for asteroid characterization approaches involve simple axis spinning models, which are not valid for tumbling debris and controlled satellites. Furthermore, asteroid inverse problems are solved in large batches where large data sets are considered with varied viewing

geometry, and frequently solutions are not possible because of lack of data. This may not be possible under the limitations of the SSN since only a relatively small number of observations may be available for each object. For SSA applications, light curve inversion approaches that can handle short arcs of observation are appealing. This motivates developing Bayesian inversion approaches for examining the light curve inversion problem for SSA applications, which is the basis of the work presented in this paper.

A number of approaches have been developed for light curve inversion for artificial satellites. Reference [11] uses light curves and thermal emissions to recover the three-dimensional shape of an object assuming its orientation with respect to the observer is known. Reference [3] uses a linear inversion model to solve for shape parameters from single-band light curve measurements. The model uses a discretization of the normal vector space to solve for albedo variants in an analogous manner to the asteroid characterization community [3]. The attitude estimation problem has also been studied assuming known shapes. Reference [12] uses an Unscented Kalman Filter approach to estimate the attitude and angular velocity with a known shape. This built a foundation for Ref. [2] which uses a bank of models with assumed shapes to perform shape and attitude estimation. The approach of Ref. [2] is effective since it marginalizes out the nonlinear shape states, and uses weights of the models to provide shape estimation. Reference [5] expands on these concepts by developing a Particle Filtering approach for estimating attitude and angular velocities for highly maneuverable objects while considering shape uncertainties. Finally, Ref. [6] uses a Marginalized Particle Filter with an exponentially correlated process noise model to estimate attitude with improved accuracy over previous approaches at a reduced computational cost over the Particle Filtering approaches used in Ref. [5].

Although much progress has been made for the light curve inversion problem with a number of attitude estimation approaches, as well as providing initial results toward full shape estimation, there is still no viable solution for the simultaneous shape, attitude and surface parameter inversion problem from light curve measurements, which is called the full inversion problem here. The full inversion problem is difficult because shape inversion from light curve measurements is an ill-posed problem, often having multiple solutions, large ambiguity, and may be non-Gaussian in nature. For example, objects with axis of symmetry clearly have attitude ambiguity for the angle about the symmetry axis. Although some ambiguities can be accounted for on a case-by-case basis, in general for the full inversion problem there is typically no *a priori* information to do this. This work proposes a Bayesian inversion method for light curve data processing to overcome these issues. Bayesian inversion methods are capable of solving for the full probability density function (pdf), which can fully resolve ambiguities, even for non-Gaussian problems. This work also formulates the light curve inversion problem in a way that makes it as flexible as possible and effective for

artificial satellites. A parameterization of light curve models is effectively used here, and definitions of prior distributions are provided to regularize these parameters, resulting in a flexible Bayesian light curve inversion approach.

One of the most general and widely used approaches for Bayesian inversion is based on Markov Chain Monte Carlo (MCMC) algorithms. These algorithms produce samples from the posterior pdf using relatively simple proposal and rejection rules. The first and very popular MCMC algorithm is the Metropolis-Hastings (MH) algorithm [13, 14], which has seen widespread uses, and has been applied to solving complicated high-dimensional statistical problems. The basic process in the MH approach starts with an initial parameter state, and uses a proposal distribution to generate proposal states, leading to the application of an acceptance-rejection criterion such that the steady-state distribution for the chain is the target distribution. This process has a number of parameters that require tuning, such as the proposal distribution and the burn-in time, i.e. the time that the chain is allowed to run to remove correlations with the initial starting point. Tuning these parameters is not trivial, and a number of adaptive approaches have been developed to estimate these parameters, and improve the overall accuracy and efficiency of the MH algorithm.

Adaptive MCMC approaches deal with the tuning problem by automatically "learning" improved parameters based on previous chain values [15]. Some of these adaptive algorithms include Adaptive Metropolis [16], Delayed Rejection (DR) [17], and Adaptive DR [18]. It has been shown that adaptive MCMC algorithms that adapt based on previous chain values will not always preserve the stationary distribution [19]. An efficient solution to the tuning problem is proposed with the Hamiltonian Monte Carlo (HMC) [20], which uses Hamiltonian dynamics to improve mixing and the overall efficiency. The HMC approach has been applied to a number of statistical applications due to a number of its useful properties. The HMC simulates a Hamiltonian system based on the target density and an auxiliary momentum variable. The HMC reduces the random walk behavior of MH, and proposes samples that may be distant from the current state. This reduction of the random walk behavior makes HMC very appealing for high-dimensional applications. However, the HMC approach still has a tuning parameter, mainly the mass matrix or the covariance of the momentum states. To overcome this Ref. [21] proposes a new method, called Riemannian Manifold HMC (RMHMC), that exploits the Riemannian geometry of the parameter space to improve the efficiency of the standard HMC. The RMHMC uses the local Fisher Information Matrix (FIM) to define the mass matrix and the scale for the momentum variable.

This work studies an Adaptive Hamiltonian Markov Chain Monte Carlo (AHMC) approach to solve the light curve Bayesian inversion problem. The shape, attitude and surface parameters of an SO are estimated using the AHMC approach, and the pdfs for these parameters are recovered. The organization of this paper is as follows. First, a description of

the problem is given. Following this, MCMC for Bayesian inversion is discussed. Then, the HMC approach is outlined, and a FIM adaptive HMC approach is given. Finally, a number of simulation test cases are shown to highlight the ability of the AHMC to determine shape parameters and recover non-Gaussian pdfs associated with the inversion problem.
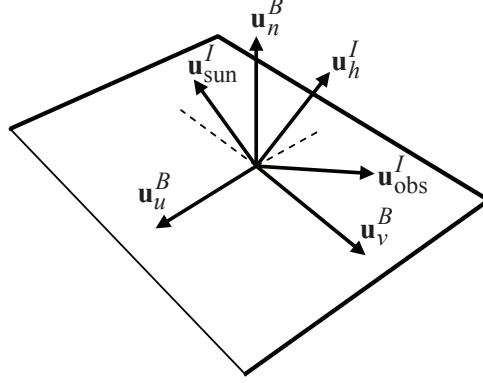

Figure 1. Reflection Geometry

II.   Light Curve Inversion Problem

The light curve inversion problem can be stated as follows: given light curve measurements, estimate dynamic states and model parameters of an SO from brightness observations $\tilde{y}_k$ for $k = 1, \ldots, m$. The dynamic states and parameters are represented by $\boldsymbol{\theta}$, and the goal is to determine $\boldsymbol{\theta}$ and its error distribution. From Figure 1, consider the following measurement relation, assuming m scalar measurements with uncorrelated measurement errors:

$$\tilde{y}_i = f(\mathbf{u}_{\text{sun}}^I(t_i), \mathbf{u}_{\text{obs}}^I(t_i), t_i, \boldsymbol{\theta}) + v_i, \quad i = 1, \ldots, m \tag{1}$$

where $v_i \sim \mathcal{N}(v_i; 0, \sigma_{m_i}^2)$, $E\{v_i v_j\} = \sigma_{m_i}^2 \delta_{ij}$, δ_{ij} is the Dirac delta function, and $\mathcal{N}(n_i; 0, \sigma_i^2)$ is a normal distribution for random variable n_i with mean 0 and variance σ_i^2. The vectors $\mathbf{u}_{\text{sun}}^I(t_i)$ and $\mathbf{u}_{\text{obs}}^I(t_i)$ denote the unit vector from the SO to the Sun and the unit vector from the SO to the observer, respectively, and together they define the observation geometry. The notation superscript I denotes that the vector is expressed in inertial coordinates. Also, t_i denotes the ith time step. Each facet has a set of three orthogonal basis vectors, $\{\mathbf{u}_n^B, \mathbf{u}_u^B, \mathbf{u}_v^B\}$, associated with it as defined in Figure 1. The notation superscript B denotes that the vector is expressed in body coordinates. The unit vector $\mathbf{u}_n^B$ points in the direction of the outward normal of the facet and the vectors $\mathbf{u}_u^B$ and $\mathbf{u}_v^B$ lay in the plane of the facet. The vector $\mathbf{u}_h^I$ is the normalized half vector between $\mathbf{u}_{\text{sun}}^I$ and $\mathbf{u}_{\text{obs}}^I$; this vector is also known as the Sun-RSO-Observer bisector. Each facet has an area $\mathcal{A}(i)$ associated with it,

where i is the facet identifier. Once the number of facets has been defined, the basis vectors $\{\mathbf{u}_n^B(i),\ \mathbf{u}_u^B(i),\ \mathbf{u}_v^B(i)\}$ and areas $\mathcal{A}(i)$ define the size and shape of the RSO.

The state and parameters may have constraints. For example, the dynamics of the SO can be viewed as a constraint on the evolution of the state. These constraints can be written as

$$g_k(\boldsymbol{\theta}) = 0, \quad k = 1, \ldots, \ell \tag{2}$$

In general, an optimal solution for the estimate of the state and model parameters is given by solving the following problem:

$$\begin{aligned}
\underset{\boldsymbol{\theta}}{\text{minimize}} \quad & \sum_{i=1}^{m} \frac{1}{2\sigma_{m_i}^2} \left\| \tilde{y}_i - f(\mathbf{u}_{\text{sun}}^I(t_i), \mathbf{u}_{\text{obs}}^I(t_i), t_i, \boldsymbol{\theta}) \right\|_2^2 \\
\text{subject to} \quad & g_k(\boldsymbol{\theta}) = 0, \quad k = 1, \ldots, \ell
\end{aligned} \tag{3}$$

The set of equations above are difficult to solve because the light curve model is in general nonlinear. Also, the state dynamics model is not known exactly because external and/or control torques may be unknown or not known precisely. Finally, the shape geometry is only an approximation since the number of parameters required to model any general shape geometry render solutions of Eq. (3) intractable.

For illustrative purposes a simple light curve model can be analyzed assuming that the object consists of a collection of N facets with only diffuse reflection, which is modeled using a non-Lambertian [22] Bidirectional Reflectance Distribution Function (BRDF) model. The energy flux as seen by an observer can be written as

$$f(\boldsymbol{\theta}, t) = \sum_{i=1}^{N} \frac{\Phi_{\text{sun,vis}}}{\pi r^2} \alpha_d(i) C \left\{ \mathbf{u}_{\text{sun}}^I(t) \cdot \left[A(t)\mathbf{u}_n^B(i) \right] \right\} \left\{ \mathbf{u}_{\text{obs}}^I(t) \cdot \left[A(t)\mathbf{u}_n^B(i) \right] \right\} \tag{4}$$

with

$$\begin{aligned}
C = & \left(1 - \left\{ 1 - \mathbf{u}_{\text{sun}}^I(t) \cdot \left[A(t)\mathbf{u}_n^B(i) \right] / 2 \right\}^5 \right) \\
& \times \left(1 - \left\{ 1 - \mathbf{u}_{\text{obs}}^I(t) \cdot \left[A(t)\mathbf{u}_n^B(i) \right] / 2 \right\}^5 \right)
\end{aligned} \tag{5}$$

where $\Phi_{\text{sun,vis}} = 455 \text{ W/m}^2$ is the solar power per unit area in the visible band, and r denotes the range between the SO and the observer. The matrix $A(t)$ denotes the mapping from body coordinates to inertial frame with kinematic relationship $\dot{A}(t) = [\boldsymbol{\omega}(t)\times] A(t)$, where $\boldsymbol{\omega}(t)$ is the angular velocity of the SO and $[\boldsymbol{\omega}(t)\times]$ is the standard cross product matrix [23]. Since the SO is assumed to have unknown shape there is no preferred body fixed frame that is known *a priori*, and therefore, without loss of generality, it can be assumed that $A(t_0) = I_{3\times3}$ where t_0 denotes the time of the first measurement. Here $\alpha_d(i)$ denotes the

diffuse albedo area product, $\alpha_d(i) = s(i)F_0(i)\mathcal{A}(i)$ (see Table 1 in the appendix). Note that since $A(t_0) = I_{3\times3}$ is assumed the attitude state can be neglected, and therefore the dynamic state of the SO only includes $\boldsymbol{\omega}(t)$. A number of estimation models exist for $\boldsymbol{\omega}(t)$, but for this work it will be assumed that $\dot{\boldsymbol{\omega}}(t) = \mathbf{0}$.

Then under the faceted model assumption, and only considering diffusion reflection, the albedo normalized shape of the SO can be described by the collection of $\alpha_d(i)$ and $\mathbf{u}_n^B(i)$. The unit vector, $\mathbf{u}_n^B(i)$, can parameterized using two parameters ϕ_i and g_i, so that

$$\mathbf{u}_n^B(i) = \begin{bmatrix} \cos(\phi_i)\sqrt{1 - g_i^2} \\ \sin(\phi_i)\sqrt{1 - g_i^2} \\ g_i \end{bmatrix} \tag{6}$$

Therefore using the measurement model described in Eq. (4), the parameter vector in Eq. (1) is given by $\boldsymbol{\theta} = [\boldsymbol{\omega}^T, \alpha_d(1), g_1, \phi_1, \cdots, \alpha_d(N), g_N, \phi_N]^T$. The albedo area and normal vector parameters are constrained, where $0 < \alpha_d(i)$, $0 < \phi_i < 2\pi$, and $-1 < g_i < 1$. These constraints can be incorporated into the prior distribution on the variables $\alpha_d(i)$, ϕ_i and g_i. This work accounts for the constraints on ϕ_i and g by defining a uniform prior distribution for these parameters given by $p(\phi_i) = U[0, 2\pi]$ and $p(g_i) = U[-1, 1]$, respectively. It is also desired that the area estimates be as small as possible, since SO areas are constrained by launch vehicle capability. An estimate that favors smaller areas is preferred. Therefore, a log-normal distribution is used for the prior on $\alpha_d(i)$, given by

$$p(\alpha_d(i)) = \frac{1}{\alpha_d(i)\sigma_\alpha\sqrt{2\pi}} \exp\left\{ \frac{-[\log(\alpha_d(i)) - \mu_\alpha]^2}{2\sigma_\alpha^2} \right\} \tag{7}$$

where σ_α is the standard deviation, and μ_α is the mean, both assumed to be given. This prior behaves like a regularization term in that it will skew the posterior distribution towards smaller areas and keep the area estimates well behaved. Finally, from Eq. (4) it can be seen that not all facets contribute to a given observation. Mainly, for a facet to contribute to a observation it must be both illuminated and visible to the observer. These two conditions are captured by the $\mathbf{u}_{\text{sun}}^I(t) \cdot \left[A(t)\mathbf{u}_n^B(i)\right]$ and $\mathbf{u}_{\text{obs}}^I(t) \cdot \left[A(t)\mathbf{u}_n^B(i)\right]$ terms, respectively.

For a given measurement vector, $\tilde{\mathbf{y}} = [\tilde{y}_1, \ldots, \tilde{y}_m]^T$, there may be some locations (ϕ_i, g_i) that are not illuminated and/or visible. Therefore the MCMC approach should not explore these parts of the state space, and a illuminated and visible prior is used such that

$$p(\phi_i, g_i) \propto \max\left[0, \text{sign}\left(\sum_k^m \left\{\mathbf{u}_{\text{sun}}^I(t_k) \cdot \left[A(t_k)\mathbf{u}_n^B(i)\right]\right\} \left\{\mathbf{u}_{\text{obs}}^I(t_k) \cdot \left[A(t_k)\mathbf{u}_n^B(i)\right]\right\}\right)\right] \tag{8}$$

where this prior is zero if the facet location, (ϕ_i, g_i), is never both illuminated and visible. The prior on the angular velocity can be defined using some notional bounds on expected angular velocities for both active and debris SOs. Then a Gaussian distribution can over-bound this expected maximum angular velocity. Here it is assumed that $p(\boldsymbol{\omega}) = \mathcal{N}(\boldsymbol{\omega}; \mathbf{0}, \sigma_\omega^2 I_{3\times3})$. Finally, the full prior of the full parameter vector is given by

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\omega}) \prod_i^N p(\phi_i, \ g_i) p(\alpha_d(i)) p(g_i) p(\phi_i) \tag{9}$$

The prior distribution does have some flexibility that can be used to control the robustness and efficiency of the Bayesian Inversion method. Measurements of the flux are corrupted with noise, modeled by $\tilde{y}_i = f(\boldsymbol{\theta}, t_i) + v_i$. Then assuming that the noise terms, v_i, are white, the likelihood function for the measurement model is given by

$$p(\tilde{\mathbf{y}}|\boldsymbol{\theta}) = \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left\{ \sum_{i=1}^m -\frac{[\tilde{y}_i - f(\boldsymbol{\theta}, t_i)]^2}{2\sigma_m^2} \right\} \tag{10}$$

where the measurement variance, σ_m, is assumed to be the same for all observations.

III.   Markov Chain Monte Carlo for Bayesian Inversion

In the Bayesian setting, the parameter vector denoted by $\boldsymbol{\theta}$ is embedded with a prior distribution $p(\boldsymbol{\theta})$. This prior is used to account for the constraints and other known properties of the parameters. For inversion problems that are ill-posed the prior distribution can be used to regularize the parameter space by enforcing a certain structure. The observation vector is given a probabilistic model based on the measurement noise statistics, where it is assumed that

$$\tilde{\mathbf{y}} \sim p(\cdot|\boldsymbol{\theta}) \tag{11}$$

where $p(\cdot|\boldsymbol{\theta})$ denotes the likelihood distribution. Then Bayesian inference is based on the posterior distribution, $p(\boldsymbol{\theta}|\tilde{\mathbf{y}})$. Bayes' Rule [24] relates the prior and likelihood distribution to the posterior distribution:

$$p(\boldsymbol{\theta}|\tilde{\mathbf{y}}) = \frac{p(\tilde{\mathbf{y}}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\tilde{\mathbf{y}}|\mathbf{s})p(\mathbf{s})d\mathbf{s}} \tag{12}$$

where the denominator is a normalization constant to make $p(\boldsymbol{\theta}|\tilde{\mathbf{y}})$ a proper pdf, and in most cases is not tractable. The normalizing constant makes computation of the posterior distribution especially difficult for multidimensional problems. The MCMC approach provides a simple method for simulating values from a distribution that can be calculated only up to a normalizing constant.

Therefore the issue with the normalizing constant is overcome by MCMC-based methods, which rather than forming $p(\boldsymbol{\theta}|\tilde{\mathbf{y}})$, produces a Markov Chain $\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \cdots, \boldsymbol{\theta}^N\}$ which is a sampling from $p(\boldsymbol{\theta}|\tilde{\mathbf{y}})$. This chain can be used to estimate moments and statistics of the posterior distribution, thereby allowing for the recovery of an estimate of the full posterior distribution. The Markov chains are generated using information from an unnormalized posterior distribution through

$$\pi(\boldsymbol{\theta}|\tilde{\mathbf{y}}) \propto p(\tilde{\mathbf{y}}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{13}$$

where $\pi(\boldsymbol{\theta}|\tilde{\mathbf{y}})$ denotes the target distribution. The algorithm used in the simulation ensures that the chain will take values in the domain of the unknown $\boldsymbol{\theta}$, and that its limiting distribution (as $N \to \infty$) will be the target distribution $\pi(\boldsymbol{\theta})$.

The MH algorithm is currently the most general algorithm for MCMC simulation. MH proposes moves in the parameter space, and accepts or rejects the moves based on the ratio of $\pi(\boldsymbol{\theta})/\pi(\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ are the current and proposed parameter vector, respectively. By calculating this ratio the integral constant in Bayes' Rule cancels out. The MH algorithm has a transition pdf given by

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*), \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}^* \tag{14a}$$

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1 - \int q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*)d\boldsymbol{\theta} \tag{14b}$$

where $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is the proposal density, $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is the acceptance probability, $p(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is the probability of moving to $\boldsymbol{\theta}^*$ from $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta}, \boldsymbol{\theta})$ is the probability of staying at $\boldsymbol{\theta}$. The density $q(\boldsymbol{\theta}, \cdot)$, with $\boldsymbol{\theta}$ being the current location of the chain, is the proposal density and used to generate random proposal states. MCMC algorithms have to satisfy the reversibility condition, which is a sufficient condition for the density $\pi(\boldsymbol{\theta})$ to be the stationary distribution of the chain. The reversibility condition is stated as

$$\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \pi(\boldsymbol{\theta})q(\boldsymbol{\theta}^*, \boldsymbol{\theta})\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}) \tag{15}$$

The stationary distribution implies

$$\int \pi(\boldsymbol{\theta})p(\boldsymbol{\theta}, \boldsymbol{\theta}^*)d\boldsymbol{\theta} = \boldsymbol{\theta}^* \tag{16}$$

This leads to the choice of the MH acceptance probability α as

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \max\left[1, \frac{\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}\right] \tag{17}$$

Then once a proposal distribution, $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$, and desired length of chain are selected the

MH-MCMC algorithm is given by

1. Choose an initial value $\boldsymbol{\theta}^0$;

2. At each step, where the current value is $\boldsymbol{\theta}_{i-1}$, propose a candidate for the new parameter $\boldsymbol{\theta}^*$ from the distribution $q(\boldsymbol{\theta}_{i-1}, \cdot)$;

3. If the proposed value $\boldsymbol{\theta}^*$ has a higher target density value than $\boldsymbol{\theta}_{i-1}$ or

$$\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > \pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$$

the proposal is accepted unconditionally;

4. If this is not the case, then $\boldsymbol{\theta}^*$ is accepted as the new value with a probability α given by Eq. (17);

5. If $\boldsymbol{\theta}^*$ is not accepted, then the chain remains at the current value and $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i-1}$;

6. Repeat the simulation from step 2 until the desired length of the chain is reached.

IV.   Hamiltonian Markov Chain Monte Carlo

The Hamiltonian or Hybrid Monte Carlo (HMC) method provides a more efficient mixing and burn-in as compared to a traditional Metropolis algorithm by avoiding the random walk behavior and proposing samples with close to one acceptance probability. The proposal system uses a Hamiltonian system, $H(\boldsymbol{\theta}, \mathbf{p})$, that includes the parameter vector $\boldsymbol{\theta}$ and an auxiliary momentum variable $\mathbf{p}$. Then proposals are generated by simulating the Hamiltonian dynamics of $H(\boldsymbol{\theta}, \mathbf{p})$ for some fictitious time. The Hamiltonian is then defined by a potential and kinetic energy function

$$H(\boldsymbol{\theta}, \mathbf{p}) = V(\boldsymbol{\theta}) + T(\boldsymbol{\theta}, \mathbf{p}) \tag{18}$$

where the potential energy, $V(\boldsymbol{\theta})$, is taken to be the negative log of the target density, $V(\boldsymbol{\theta}) = -\log(\pi(\tilde{\mathbf{y}}|\boldsymbol{\theta}))$. Choosing the potential as $-\log(\pi(\tilde{\mathbf{y}}|\boldsymbol{\theta}))$ will drive the Hamiltonian towards a high probability state by providing a gradient decent term. A clear analogy between the Hamiltonian, $H(\boldsymbol{\theta}, \mathbf{p})$, and the joint distribution $p(\boldsymbol{\theta}, \mathbf{p})$ is given by $p(\boldsymbol{\theta}, \mathbf{p}) \propto \exp(-H(\boldsymbol{\theta}, \mathbf{p}))$. Then the momentum state distribution is related to the kinetic energy function $T(\boldsymbol{\theta}, \mathbf{p})$. Since the auxiliary momentum variable is not included in the parameter vector, a random momentum vector is generated before each proposal and before simulating the Hamiltonian system. For example, it can be assumed that the kinetic energy does not depend on $\boldsymbol{\theta}$ and

has the form

$$T(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} \tag{19}$$

Then the initial momentum is randomly sampled from $p(\mathbf{p}) \propto \exp(-T(\mathbf{p}))$ or $\mathbf{p} \sim \mathcal{N}(\mathbf{0}, M)$. Then the Hamiltonian is given by

$$H(\boldsymbol{\theta}, \mathbf{p}) = V(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} \tag{20}$$

The trajectories that follow the Hamiltonian dynamics are given by

$$\dot{\boldsymbol{\theta}} = \frac{\partial H}{\partial \mathbf{p}} = M^{-1}\mathbf{p} \tag{21a}$$

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = -\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) \tag{21b}$$

The solution flow for the differential equations above, $(\boldsymbol{\theta}(\tau), \mathbf{p}(\tau)) = \varphi(\boldsymbol{\theta}(0), \mathbf{p}(0))$, preserve the total energy, i.e. $H(\boldsymbol{\theta}(\tau), \mathbf{p}(\tau)) = H(\boldsymbol{\theta}(0), \mathbf{p}(0))$ and hence the joint density $p(\boldsymbol{\theta}(\tau), \mathbf{p}(\tau)) = p(\boldsymbol{\theta}(0), \mathbf{p}(0))$. They also preserve the volume element $d\boldsymbol{\theta}(\tau)d\mathbf{p}(\tau) = d\boldsymbol{\theta}(0), d\mathbf{p}(0)$, and are time reversible. To generate a proposal state Eq. (21) must be numerically solved for some integration time t. The integration time is arbitrary and chosen such that the Hamiltonian dynamics is allowed to evolve into a higher probability state. The joint probability of the new state is given by $p(\boldsymbol{\theta}^*, \mathbf{p}^*) \propto \exp(-H(\boldsymbol{\theta}^*, \mathbf{p}^*))$. Then the acceptance probability is given by

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \max\left[1, \exp\left\{-H(\boldsymbol{\theta}^*, \mathbf{p}^*) + H(\boldsymbol{\theta}, \mathbf{p})\right\}\right] \tag{22}$$

Since the Hamiltonian is preserved $H(\boldsymbol{\theta}^*, \mathbf{p}^*) = H(\boldsymbol{\theta}, \mathbf{p})$, at least approximately, and therefore proposals generated with the HMC have nearly 1 acceptance probability. This property makes the HMC approach very efficient. There are a class of numerical integrators for Hamiltonian systems which preserve energy, volume, and are time reversible. The Stormer-Verlet or Leapfrog methods are from this class. The Leapfrog method will be used for this work, and Eq. (21) can be evolved forward by ϵ with the following:

$$\mathbf{p}(\tau + \epsilon/2) = \mathbf{p}(\tau) - \epsilon\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}(\tau))/2 \tag{23a}$$

$$\boldsymbol{\theta}(\tau + \epsilon) = \boldsymbol{\theta}(\tau) + \epsilon M^{-1}\mathbf{p}(\tau + \epsilon/2) \tag{23b}$$

$$\mathbf{p}(\tau + \epsilon) = \mathbf{p}(\tau + \epsilon/2) - \epsilon\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}(\tau + \epsilon))/2 \tag{23c}$$

The equations above only approximately conserve the total energy, introducing errors due to the numerical approximation which creates a bias in the joint density, $p(\boldsymbol{\theta}, \mathbf{p})$. This bias is corrected by the accept-reject step used in MCMC approaches. Therefore, if the numerical bias introduced causes the trajectory to head to a decreased probability state

then it will likely be rejected. But if the trajectory creates an increase in probability then it will be accepted with probability 1. If the integration error in the total energy is small, then the acceptance probability will remain at a high level. In general, since the acceptance probability for HMC is nearly one, its sampling from the invariant density $p(\boldsymbol{\theta}|\mathbf{y})$ can be considered as a Gibbs sampler where the momentum $\mathbf{p}$ acts simply as an auxiliary variable. Equations (23a) and (23b) can be written as

$$\boldsymbol{\theta}(\tau + \epsilon) = \boldsymbol{\theta}(\tau) - \frac{\epsilon^2}{2} M^{-1} \nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}(\tau)) + \epsilon M^{-1} \mathbf{p}(\tau) \tag{24}$$

which can be viewed as a pre-conditioned Langevin diffusion or gradient decent. It is clear that the mass matrix M is critical for the performance of the HMC. The tuning of the mass matrix M may be difficult without knowledge of the target density.

V.   Adaptation in Hamiltonian Markov Chain Monte Carlo

The Riemann Manifold Hamiltonian Monte Carlo (RMHC) overcomes the difficulty of tuning the mass matrix by using the local structure of the target density when proposing. The RMHC method chooses the mass matrix to be the FIM of the joint posterior, $p(\boldsymbol{\theta}, \mathbf{p}|\tilde{\mathbf{y}})$. Therefore the mass matrix is state dependent and given by $M(\boldsymbol{\theta}) = F$, where F is given by

$$F(\boldsymbol{\theta}) = E \left\{ \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log(p(\boldsymbol{\theta}|\tilde{\mathbf{y}})) \right] \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log(p(\boldsymbol{\theta}|\tilde{\mathbf{y}})) \right]^T \right\} \tag{25}$$

However, the mass matrix now becomes state dependent, so the Hamiltonian system is no longer separable and the Leapfrog method is not applicable. Instead, a generalized Leapfrog algorithm is used to simulate non-separable dynamics generated with the FIM mass matrix. The Hamiltonian for the RMHC approach is given by

$$H(\boldsymbol{\theta}, \mathbf{p}) = V(\boldsymbol{\theta}) + \frac{1}{2} \log \left[\det \left(F(\boldsymbol{\theta}) \right) \right] + \frac{1}{2} \mathbf{p}^T F(\boldsymbol{\theta})^{-1} \mathbf{p} \tag{26}$$

To solve for the HMC on the Riemmanian manifold defined by the geometry of the parameter space the system in Eq. (26) must be solved. But this can be challenging since the system is non-separable and Hamilton's equations will require gradients of $F(\boldsymbol{\theta})$. Therefore the Adaptive HMC (AHMC) approach is used here which holds $F(\boldsymbol{\theta})$ constant during the simulation of the Hamiltonian for generation of the proposal states. Although $F(\boldsymbol{\theta})$ is not allowed to vary while solving Hamilton's equation, it is updated for each new proposal step. In effect, this uses the Riemmanian manifold geometry but doesn't require the generalized Leapfrog or further derivatives of $-\log(p(\boldsymbol{\theta}|\tilde{\mathbf{y}}))$. Then the mass matrix is set to $M = F(\boldsymbol{\theta}_{i-1})$ for

the $i - 1$ proposal step, and Eq. (23) is used to simulate the Hamiltonian dynamics for the following Hamiltonian:

$$H(\boldsymbol{\theta}, \mathbf{p}) = V(\boldsymbol{\theta}) + \frac{1}{2}\mathbf{p}^T F(\boldsymbol{\theta}_{i-1})^{-1}\mathbf{p} \tag{27}$$

Note that the term $\frac{1}{2}\log\left[\det\left(F(\boldsymbol{\theta})\right)\right]$ is no longer required since the mass matrix is held constant for the proposal step. Additionally, the partial derivatives from Eq. (25) and FIM for attitude states are calculated for the Ashikhmin-Shirley [25] light curve model in Ref. [26]. Although Ref. [26] only considers the attitude when calculating the partial derivatives, a similar procedure can be applied for the state vector considered in this work. For the model given by Eqs. (4) and (5), and the likelihood function given by Eq. (10), the FIM can be calculated by

$$F(\boldsymbol{\theta}) = \frac{1}{\sigma_m^2}\sum_{i=1}^{m}\left[\frac{\partial}{\partial\boldsymbol{\theta}}f(\boldsymbol{\theta}, t_i)\right]\left[\frac{\partial}{\partial\boldsymbol{\theta}}f(\boldsymbol{\theta}, t_i)\right]^T \tag{28}$$

where the form of the FIM will depend on the contents of the state vector, $\boldsymbol{\theta}$, and the light curve model used.

VI.   Model Selection and Sparse Solutions

The light curve models discussed earlier require the specification of the number of facet normals, N, to define the shape and surface property parameter vector. For a given N the HMC approach can be used to sample for the target density describing the *a posteriori* on the state vector $\boldsymbol{\theta}$. For most cases the true number of facets is not known, and therefore this becomes a free parameter in the inversion process. This parameter in effect defines the model to be used, and dimension of the problem increases linearly with N. The dimensionally of the light curve inversion problem is given by $d = n_{sp}N$ where n_{sp} is the number of parameters for each facet. The selection of the number of facet normals may be a difficult problem since it is desirable to use a model with minimum complexity, and one that can also approximate the data well. It is expected that artificial SOs should be well approximated by flat facets. Therefore, a relatively low number of facet normals should represent the data well. It is well known that the inversion problem can be regularized using the L_1-norm to produce sparse solutions. Consider the unconstrained version of the inverse problem described by Eq. (3). A regularization term can be added to this minimization problem to promote sparse solutions. Then the L_1-norm regularized version of the inverse problem described by Eq. (3) is given by

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \sum_{i=1}^{m}\left\{\frac{1}{2\sigma_{m_i}^2}\left\|\tilde{y}_i - f(\mathbf{u}_{\text{sun}}^I(t_i), \mathbf{u}_{\text{obs}}^I(t_i), t_i, \boldsymbol{\theta})\right\|_2^2 + \frac{\|\boldsymbol{\alpha}_d\|_1^2}{2\lambda^2}\right\} \tag{29}$$

where $\boldsymbol{\alpha}_d = [\alpha_d(1), \cdots, \alpha_d(N)]^T$ and $\|\mathbf{a}\|_1$ denotes the L_1-norm of the vector $\mathbf{a}$. The parameter λ controls the weighting between reducing the measurement residual error and enforcing that $\boldsymbol{\alpha}_d$ is sparse. This parameter can introduce a bias in the estimated parameter but can also reduce the variance of the estimation error. In fact, the bias increases with increasing λ while the variance of the estimation error decreases with increasing λ, which is similar to ridge estimation [24]. For linear inverse problems the minimization problem given in Eq. (29) can be solved using the LASSO method [27]. The LASSO method provides a solution for the parameter vector $\boldsymbol{\theta}$ as a function of λ. This L_1-norm regularization can be imposed in a Bayesian framework by considering an L_1-norm prior on the area states. Then an alternative prior can be imposed on the vector albedo areas to promote sparse solutions. To promote sparsity in the solution an L_1-norm prior is used [28] as an alternative to Eq. (3). The L_1 norm prior is given by

$$p(\boldsymbol{\alpha}) \propto \exp\left\{-\frac{\|\boldsymbol{\alpha}_d\|_1^2}{\lambda^2}\right\} \tag{30}$$

Then the Bayesian version of Eq. (29) can be written in terms of the L_1-norm prior:

$$p(\boldsymbol{\theta}|\tilde{\mathbf{y}}) \propto \exp\left\{\frac{-\|\tilde{y}_i - f(\boldsymbol{\theta}, t_i)\|_2^2}{2\sigma_{m_i}^2}\right\} \exp\left\{-\frac{\|\boldsymbol{\alpha}_d\|_1^2}{2\lambda^2}\right\} \prod_i^N p(\phi_i, g_i)p(\alpha_d(i))p(g_i)p(\phi_i)$$

Samples from $p(\boldsymbol{\theta}|\tilde{\mathbf{y}})$ are then generated using the HMC approach for a given value of λ. Note that λ is a user defined parameter, and for each given problem there is a tradeoff between biases and variance reduction when selecting λ. For this work the L_1-norm prior is used for model selection, not for determining the final Bayesian inversion state estimate. During the burn-in phase of the HMC, sampling is used for a given value of λ to find a sparse model representation. The chain generated using Eq. (30) will drive as many albedo area values to zero as possible while adjusting the remaining albedo areas to fit the data. Since the samples generated from Eq. (30) are biases due to λ, they are only used to burn-in and find a reduced model representation and a reasonable estimate of the location of the steady-state distribution. Then the final sample from this burn-in phase is used to estimate the number facets to be used in the full HMC run. The number of facets with albedo areas above a threshold will then be considered as an estimate of the number of facets needed. Given these estimates this reduced vector is used as the starting location for another HMC run to produce a sample from the target density. The process begins with running the HMC with initial value of N_i which is selected as a upper bound on the true number of facets, such that $N_i > N_t$ where N_t is the true number of facets. Then the final location of the chain from the L_1-norm regularized target density given in Eq. (30) is used to estimate N as the number of albedo areas which are higher than some lower bound, α_d^L. Finally, using the prior in Eq. (3) a chain is generated which should have a reduced bias and a good estimate
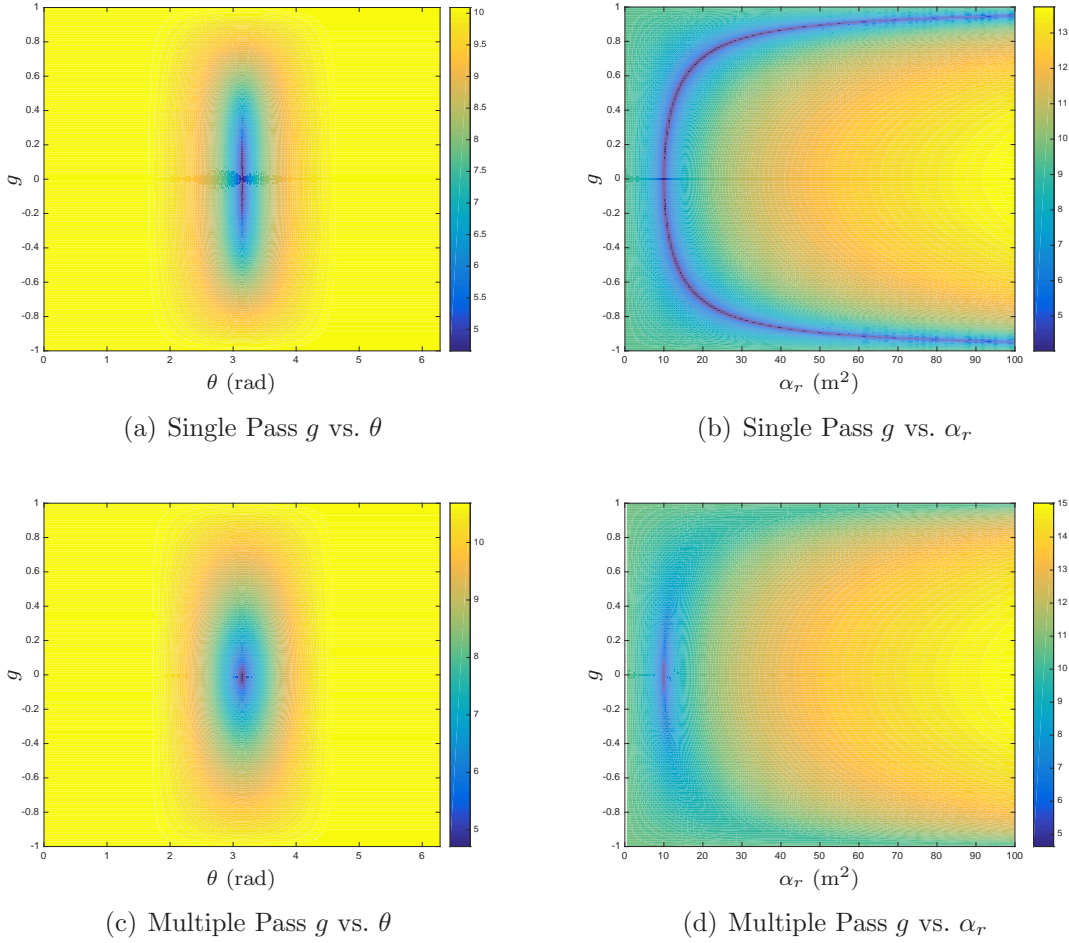
of the number of facets.

(a) Single Pass g vs. θ

(b) Single Pass g vs. α_r

(c) Multiple Pass g vs. θ

(d) Multiple Pass g vs. α_r

Figure 2. Contour of $\log\left(\log \pi(\tilde{\mathbf{y}}|\boldsymbol{\theta})\right)$

VII. Simulation Results

Several simulation test cases are considered to analyze the performance of the AHMC approach. First the target distribution is studied for a single facet model using a non-Lambertian diffuse BRDF shown in Table 1 in the appendix. Then the HMC approach is applied for approximating a draw from the target distribution for the single facet model. The pdf for the chain is calculated along with the predicted measurement output. Next, the multiple facet case is considered with a six facet truth model. For this case the HMC approach correctly assumes a state vector with six facets. Again the non-Lambertian diffuse BRDF model is used for this case. All cases use a simple measurement geometry for illustrative purposes, where the illumination and observer geometry used to generate measurements are simplified. These examples consist of relatively simple measurement geometry where the

observer and Sun directions are defined as

$$
\begin{aligned}
\mathbf{u}_{\mathrm{obs}}^{I}(t_i) &= [\cos(\omega_{\mathrm{obs}}t_i + \lambda_{\mathrm{obs}}), \quad \sin(\omega_{\mathrm{obs}}t_i + \lambda_{\mathrm{obs}}), \quad 0]^{T} \\
\mathbf{u}_{\mathrm{sun}}^{I}(t_i) &= [\cos(\omega_{\mathrm{sun}}t_i + \lambda_{\mathrm{sun}}), \quad \sin(\omega_{\mathrm{sun}}t_i + \lambda_{\mathrm{sun}}), \quad 0]^{T}
\end{aligned}
\tag{31}
$$

where t_i is a time parameter, and ω_{obs} and ω_{sun} are the angular velocity of the Sun and observer directions, respectively. In general, these directions can be defined in a realistic manner by simulating the SO orbit and determining the directions to a ground site and the Sun given the SO orbital position. Since it is desired to study the Bayesian light curve inversion problem, it is beneficial to forgo a complete orbital simulation for the model shown in Eq. (31). Then by defining ω_{obs}, ω_{sun}, λ_{obs}, and λ_{sun} the simulation geometry can be completely specified.
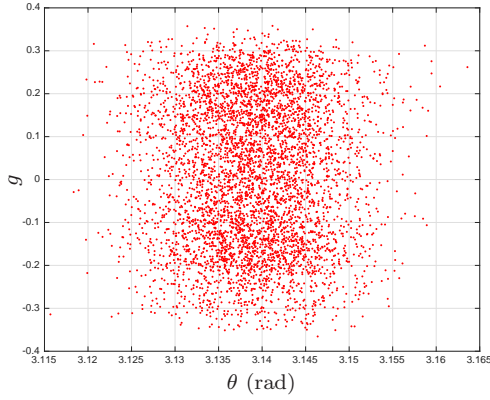
(a) Single Pass g vs. θ
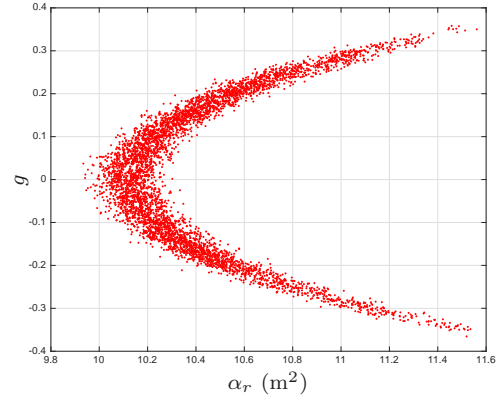
(b) Single Pass g vs. α_r
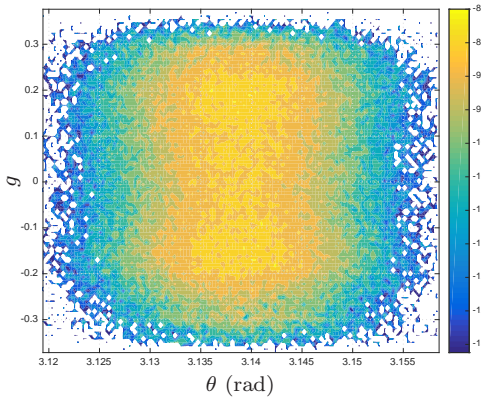
(c) Multiple Pass g vs. θ
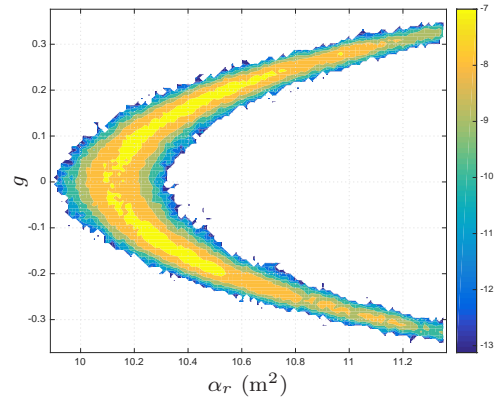
(d) Multiple Pass g vs. α_r

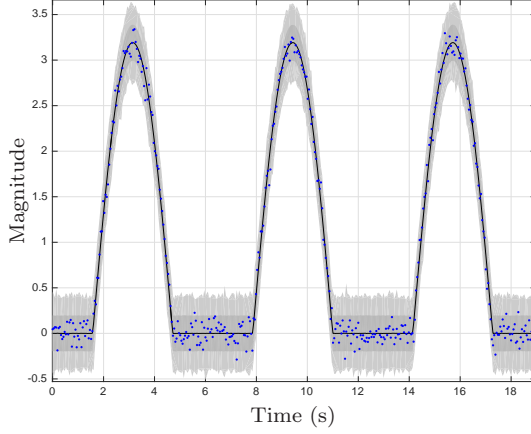**Figure 3.** Contour of $\log\left(\pi(\tilde{\mathbf{y}}|\boldsymbol{\theta})\right)$

Figure 4. Predicted Measurements

A. Single Non-Lambertian Diffuse Facet Target Density

In this section the target density for a single facet model is analyzed. Flux measurements are generated for a single facet model using the non-Lambertian diffuse model given in Table 1. The measurement likelihood and prior pdf given by Eqs. (10) and (9) are used to calculate the target distribution given by Eq. (13). The single facet model with non-Lambertian diffuse BRDF consists of a parameter vector given by $\boldsymbol{\theta} = [\alpha_r, \ \theta, \ g]^T$. Then two measurement geometries are considered, one with a single pass direction using Eq. (31) and one case with two passes with orthogonal trajectories ($\mathbf{u}_{\text{obs}}^I(t_i)$ on the second pass is not on the y-z plane). Then the target density is evaluated at different points in the parameter space, which is shown in Figure 2. Figures 2(a) and 2(b) show the loglog of target density for a single pass. Figure 2(b) shows the target density in the θ-g space. From this figure it can be seen that the location of the facet is weakly observable in the direction perpendicular to the pass. The location has high variation in the g direction and smaller variation in the θ direction. Figure 2(a) shows the target density in the θ-a space. Here it can be seen the target density is highly non-Gaussian. From this figure it is clear that for a true location of $g = 0$, the measurement can be reproduced by high value of g with a larger area. Increasing g results in a larger angle between the observer and the normal, but by increasing the area the effect of increasing this quantity can be negated. Figures 2(c) and 2(d) show the target density plots for two passes. In this case the observer direction is simulated using Eq. (31) with motion in the x-y plane and also with motion in the y-z plane. The multiple passes reduce the uncertainty in the g direction as seen by Figure 2(d), and the area is also further contained as seen by Figure 2(c).
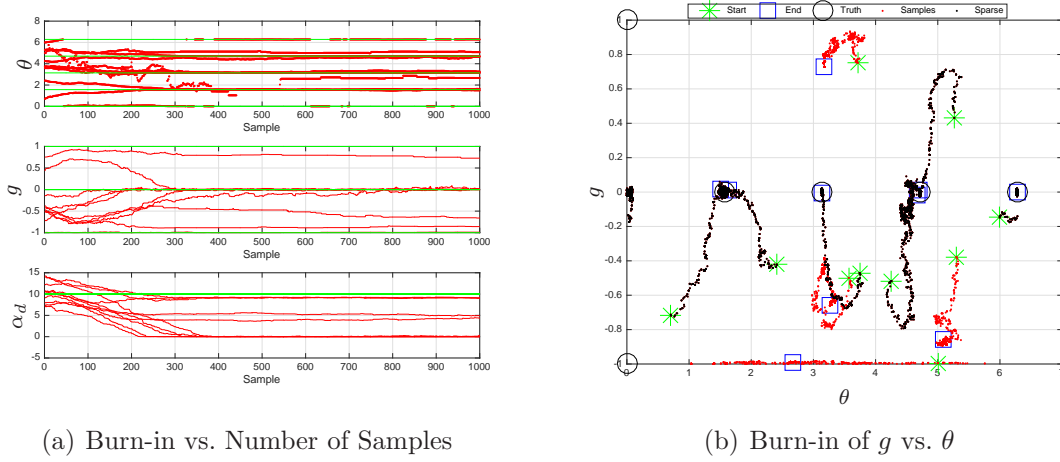
(a) Burn-in vs. Number of Samples

(b) Burn-in of g vs. θ

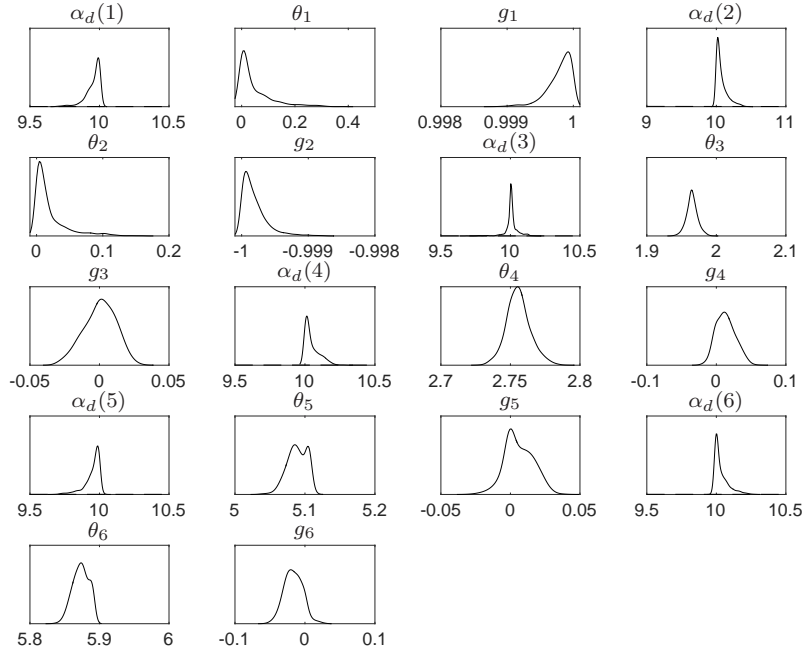**Figure 5.** $L1$ **Regularized HMC Results**

Figure 6. **Empirical Probability Density Function for Multiple Facet Model**

B.   AHMC Single Facet

The AHMC approach is applied to the single facet inversion problem. In this case one single measurement pass is considered using Eq. (31). The AHMC approach is used to generate proposal states, and the MH-MCMC acceptance condition is used to accept or reject the proposal. The initial chain location is sampled from the prior density given by Eq. (9). The chain is allowed to burn-in for 1,000 samples so that the memory of the starting location is
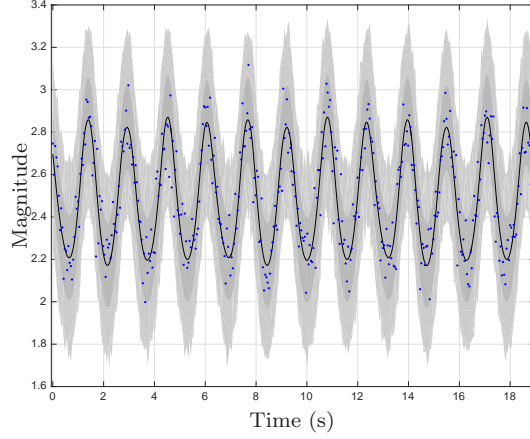
Figure 7. Predicted Measurements for Multiple Facet Example

removed. Figure 3 shows the result of the single facet AHMC runs. The pdf for the chain is estimated using the samples, which is shown in Figures 3(c) and 3(d). The AHMC samples are shown in Figures 3(a) and 3(b). From these figures it can be seen that the samples have the shape of the target density that has been seen in Figure 2. The pdf is estimated from the sample using a histogram. The estimated pdf is shown in Figures 3(b) and 3(a). From this figure it is seen that the single facet model is non-Gaussian, and the AHMC chain closely approximates the target density as seen by Figure 2.

C.   AHMC Multiple Facet

For this case the multiple facet model is considered, where the measurement is simulated for a six facet model with truth values given by

$$\boldsymbol{\alpha}_d = [10,\ 10,\ 10,\ 10,\ 10,\ 10]^T \tag{32a}$$

$$\boldsymbol{\theta} = [0,\ 0,\ \pi/2,\ \pi,\ 3\pi/2,\ 2\pi]^T \tag{32b}$$

$$\mathbf{g} = [1,\ -1,\ 0,\ 0,\ 0,\ 0]^T \tag{32c}$$

The standard deviation of measurement uncertainty is given by $\sigma_m = 0.1$, and flux measurements are generated using the single pass defined by Eq. (31). Again the initial chain location is randomly sampled from the prior density, and the chain is allowed to burn-in for 1,000 samples. The burn-in phase uses the sparse prior discussed earlier with a fairly large sparsity promoting parameter where $\lambda = 1000$. The value of λ is determined from experimentation. For this value it is found that the AHMC approach correctly determines the true number of facets during burn-in. In practice the true number of facets is not known. An alternative approach could involve running the burn-in a number of times on the real

data until an acceptable measurement residual is achieved while promoting sparsity. After the burn-in phase the final states of the burn-in are fed into another run of the AHMC to produce samples from the target density. The burn-in phase is shown in Figure 5. The AHMC chain versus the number of samples is shown in Figure 5(a). This figure shows the initial burn-in converging to the target density. It can be seen that some of the albedo areas converge to zero while others do not. The burn-in of the facet locations is shown in Figure 5(b), where normal vector directions are adjusted to fit the data with a sparse model. The final locations of the normal vector during burn-in are close to the true locations, as seen by Figure 5(a). Figure 6 shows the empirical pdf for the multiple facet inversion problem. From this figure it can be seen that the AHMC approach can recover the non-Gaussian pdf for all of the SO shape states. The relative high dimensionality of the state space is handled well by the AHMC, and the approach shows efficient mixing in that a low number of samples is required to provide good estimates of the target density. Finally, from Figure 7 it can be seen that the chain provides a good prediction of the measurements with non-Gaussian error estimates.
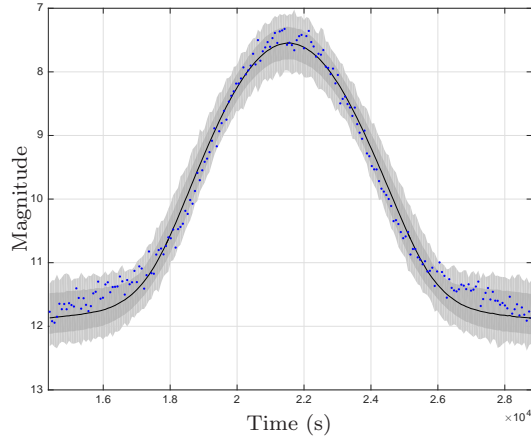

Figure 8. Geosynchronous Satellite Glint Example

D.   Geosynchronous Satellite Glint Example

This example considers light curve measurements from a Geosynchronous (GEO) satellite with stabilizing attitude control. GEO satellite glints are a well known light curve attribute, which are observed for GEO satellites with Sun tracking solar panels [29]. GEO satellites with Sun tracking solar panels will have predictable glint when the observer directions aligns with the Sun direction, and large solar panes reflect in a mostly specular manner. Since modeling of this specular reflection is important for the GEO case, the Ashikhmin-Shirley is used for both simulating the observations and as the model used in the Bayesian in-

version. The Ashikhmin-Shirley model has the following parameter vector for each facet $\boldsymbol{\theta} = [\boldsymbol{\omega}, \; \mathcal{A}(i), \; \theta, \; g, \; \rho, \; n, \; d]^T$. Reference [8] uses this model to estimate the rotational and surface parameter states ρ, n and d. This example is the first time that the full shape is recoverable using the Ashikhmin-Shirley in a full Bayesian setting. The observation for this example along with the predicted observation and uncertainty are shown in Figure 8. The AHMC is again burned-in for 1,000 samples with a 10 facet model, and the sparse state vector determined from burn-in has 3 facets. From Figure 8 it can been seen that the Ashikhmin-Shirley has a strong specular reflection component. The AHMC is able to recover this specular compute and approximate the shape of the measured light curve with good performance.

VIII.   Conclusions

This work studied the space object shape and parameter inversion problem using light curve measurements. This work employs an Adaptive Hamiltonian Markov Chain Monte Carlo (AHMC) estimation approach to produce samples from the posterior probability distribution of the shape and parameter states. The target density was studied, and it was shown that for a single facet shape model the location of the facet is weakly unobservable in the direction orthogonal to the path of the observer unit vector. Also the area state has a non-Gaussian pdf, making this state weakly unobservable. But with the addition of a second pass both the area and location of the facet become more observable. The multiple facet estimation problem was also studied, where the dimension of the state is 30 during the burn-in phase. Using an L_1-norm prior during burn-in, a sparse shape model was recovered that represented the data well and reduced the number of states to 18. The AHMC performs well in the high-dimensional state-space. The pdf for the large dimensional state-space was generated and the measurement were predicted with reasonable accuracy. Finally, the AHMC approach was shown to be able to solve the Bayesian inversion problem using light curve measurements.

Appendix: Bidirectional Reflectance Distribution Functions

A. Partial Derivatives of Non-Lambertian Diffuse Model

The partial derivatives of the simplified measurement function are calculated as

$$\frac{\partial}{\partial \alpha_d(i)} f(\boldsymbol{\theta}, t_i) = \frac{zq}{\pi} C \tag{A.1a}$$

$$\frac{\partial}{\partial \phi_i} f(\boldsymbol{\theta}, t_i) = C_1 \mathbf{u}_{\text{obs}}^I(t) \left[\frac{\partial}{\partial \phi_i} \mathbf{u}_n^B(i) \right] + C_2 \mathbf{u}_{\text{sun}}^I(t) \left[\frac{\partial}{\partial \phi_i} \mathbf{u}_n^B(i) \right] \tag{A.1b}$$

$$\frac{\partial}{\partial g_i} f(\boldsymbol{\theta}, t_i) = C_1 \mathbf{u}_{\text{obs}}^I(t) \left[\frac{\partial}{\partial g_i} \mathbf{u}_n^B(i) \right] + C_2 \mathbf{u}_{\text{sun}}^I(t) \left[\frac{\partial}{\partial g_i} \mathbf{u}_n^B(i) \right] \tag{A.1c}$$

$$C_1 = \frac{\alpha_d(i)}{\pi} \left[\frac{5}{2} zq \left(1 - (1 - z/2)^4 \right) \left(1 - (1 - q/2)^5 \right) + Cq \right] \tag{A.1d}$$

$$C_2 = \frac{\alpha_d(i)}{\pi} \left[\frac{5}{2} zq \left(1 - (1 - z/2)^5 \right) \left(1 - (1 - q/2)^4 \right) + Cz \right] \tag{A.1e}$$

$$\left[\frac{\partial}{\partial \phi_i} \mathbf{u}_n^B(i) \right] = \left[-g_i \sin(\phi_i) \sqrt{1 - g_i^2}, \quad \cos(\phi_i) \sqrt{1 - g_i^2}, \quad 0 \right]^T \tag{A.1f}$$

$$\left[\frac{\partial}{\partial g_i} \mathbf{u}_n^B(i) \right] = \left[-g_i \cos(\phi_i) \left(1 - g_i^2 \right)^{-1/2}, \quad -g_i \sin(\phi_i) \left(1 - g_i^2 \right)^{-1/2}, \quad 1 \right]^T \tag{A.1g}$$

where z and q are given by $q = \mathbf{u}_{\text{sun}}^I(t) \cdot \left[A(t) \mathbf{u}_n^B(i) \right]$ and $z = \mathbf{u}_{\text{obs}}^I(t) \cdot \left[A(t) \mathbf{u}_n^B(i) \right]$, respectively.

B. Review of Bidirectional Distribution Functions

A number of BRDFs exist in the literature. These models are based on the BRDF that models light distribution scattered from the surface due to the incident light. The BRDF at any point on the surface is a function of two directions, the direction from which the light source originates and the direction from which the scattered light leaves the observed the surface. These models include Ashikhmin-Shirley [25], a simplified Blinn-Phong [30], and Cook-Torrance [31] to list some. The BRDF models how light energy reflects off of surfaces and how this reflected energy is distributed in various directions. The brightness of an object in space can be modeled using an anisotropic Phong light diffusion model [25].

The model in Ref. [25] decomposes the BRDF into a specular component and a diffuse component. The two terms sum to give the total BRDF:

$$\rho_{\text{total}}(i) = \rho_{\text{spec}}(i) + \rho_{\text{diff}}(i) \tag{A.2}$$

The diffuse component represents light that is scattered equally in all directions (Lambertian) and the specular component represents light that is concentrated about some direction (mirror-like). Reference [25] develops a model for continuous arbitrary surfaces but simplifies

for flat surfaces. This simplified model is employed in this work as shape models are considered to consist of a finite number of flat facets. Therefore, the total observed brightness of an object becomes the sum of the contribution from each facet:

$$f_r = (dR_d + sR_s) \tag{A.3}$$

which depends on the diffuse bidirectional reflectance, R_d, the specular bidirectional reflectance, R_s, and the fraction of each to the total (d and s, respectively, where $d + s = 1$). These bidirectional reflectances are calculated differently for the various models. In each model, however, $c = \mathbf{u}_{\text{obs}}^I \cdot \mathbf{u}_h^I$, ρ is the diffuse reflectance ($0 \leq \rho \leq 1$), and F_0 is the specular reflectance of the surface at normal incidence ($0 \leq F_0 \leq 1$). To be used as a prediction tool for brightness and radiation pressure calculations, an important aspect of the BRDF is energy conservation. For energy to be conserved, the integral of the BRDF times $\cos(\theta_r)$ over all solid angles in the hemisphere with $\theta_r \leq \pi/2$ needs to be less than unity, with

$$\int_0^{2\pi} \int_0^{\pi/2} f_r \cos(\theta_r) \sin(\theta_r) \, d\theta_r d\phi = R_d + R_s \tag{A.4}$$

For the BRDF given in Eq. (A.3), this corresponds to constant values of $R_d = d\rho$ and $R_s = sF_0$. The remaining energy not reflected by the surface is either transmitted or absorbed. In this paper it is assumed that the transmitted energy is zero. A review of the various BRDS functions is provided in Table 1.

Ashikhmin-Shirley BRDF: In addition to d, ρ, and F_0, the Ashikhmin-Shirley BRDF has two exponential factors (n_u, n_v) that define the reflectance properties of each surface. The Ashikhmin-Shirley diffuse and specular reflectivities are not constant, however, but rather complicated functions of illumination angle, exponential factor, and the diffuse and specular reflectances. In all cases, however, $R_d + R_s \leq 1$, so energy is conserved.

Blinn-Phong: The specular bidirectional reflectance of the original Phong model is proportional to $\left[\mathbf{u}_n^I(i) \cdot \mathbf{R}\right]^n$, where $\mathbf{R}$ is the perfect mirror-like reflection of $\mathbf{u}_{\text{sun}}^I$. Blinn [30] proposes that $\mathbf{u}_h^I$ be used instead of $\mathbf{R}$ to make it easier and faster to calculate. Unfortunately, both versions of the model do not conserve energy and thus are unsuited for the purposes of brightness estimation. The model can be made to conserve energy, however, by modifying the leading term. In keeping with the desire for simplicity in this model, the leading term is chosen to only be a function of the exponential factor and set to yield a reflectivity equal to the mirror-like reflection of Eq. (3) at normal illumination. In addition to d, ρ and F_0, the simplified Blinn-Phong BRDF has a single exponential factor, n, that defines the reflectance properties of each surface.

Cook-Torrance BRDF: This model has the facet slope distribution function, D, the

Table 1. Review of Bidirectional Distribution Functions

Lambertian	$$R_d = \frac{\rho d}{\pi}$$ $$R_s = 0$$
Non-Lambertian Diffuse	$$R_d = F_0 \left\{ 1 - \left[1 - \mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{sun}}^I/2 \right]^5 \right\}$$ $$\times \left\{ 1 - \left[1 - \mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{obs}}^I/2 \right]^5 \right\}$$ $$R_s = 0$$
Ashikhmin-Shirley	$$F = F_0 + \left(\tfrac{1}{s} - F_0 \right)(1 - c)^5$$ $$R_d = \frac{28\rho}{23\pi}(1 - sF_0) \left\{ 1 - \left[1 - \frac{\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{sun}}^I}{2} \right]^5 \right\}$$ $$\times \left\{ 1 - \left[1 - \frac{\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{obs}}^I}{2} \right]^5 \right\}$$ $$R_s = \frac{\sqrt{(n_u+1)(n_v+1)}}{8\pi} \frac{F}{c \max\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{sun}}^I, \mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{obs}}^I \right]}$$ $$\times \left[\cos(\psi) \right]^{n_u \cos^2(\beta) + n_v \sin^2(\beta)}$$ $$\cos(\psi) = \mathbf{u}_h^I \cdot \mathbf{u}_n^I(i), \quad \cos(\beta) = \mathbf{u}_h^I \cdot \mathbf{u}_v^I$$
Cook-Torrance	$$R_d = \rho/\pi$$ $$D = \frac{1}{\pi m^2 \cos^4 \psi} e^{\tan(\psi)/m^2}$$ $$G = \min \left\{ 1, \frac{2\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_h^I\right]\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{obs}}^I\right]}{\mathbf{u}_{\text{obs}}^I \cdot \mathbf{u}_h^I}, \frac{2\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{sun}}^I\right]\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_h^I\right]}{\mathbf{u}_{\text{obs}}^I \cdot \mathbf{u}_h^I} \right\}$$ $$F = \frac{(g-c)^2}{2(g+c)^2} \left\{ 1 + \frac{[c(g+c)-1]^2}{[c(g-c)+1]^2} \right\}$$ $$R_s = \frac{DGF}{4\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{sun}}^I\right]\left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_{\text{obs}}^I\right]}$$
Blinn-Phong	$$R_d = \frac{\rho d}{\pi}$$ $$R_s = (1 - d) \left[\frac{F_0(n+2)(n+4)}{8\pi(n+2^{-n/2})} \right] \left[\mathbf{u}_n^I(i) \cdot \mathbf{u}_h^I \right]$$

geometrical attenuation factor, G, and the reflectance of a perfectly smooth surface, F, with $g = n^2 + c^2 - 1$ and the index of refraction $n = \frac{1 + \sqrt{F_0}}{1 - \sqrt{F_0}}$. In addition to d, ρ, and F_0, the Cook-Torrance BRDF has a facet slope, m, parameter that defines the reflectance properties of each surface. The facet slope parameter of the Cook-Torrance BRDF and the exponential factor of the Ashikhmin-Shirley and Blinn-Phong BRDFs are roughly related by $n = 2/m^2$.

References

[1] House, W., "National Space Policy of the United States of America," *Retreived from https://www.whitehouse.gov/sites/default/files/national_space_policy_6–28–10.pdf* , 2010.

[2] Linares, R., Jah, M. K., Crassidis, J. L., and Nebelecky, C. K., "Space Object Shape Characterization and Tracking Using Light Curve and Angles Data," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 1, Jan.-Feb. 2013, pp. 13–25, doi:10.2514/1.62986.

[3] Hall, D., Calef, B., Knox, K., Bolden, M., and Kervin, P., "Separating Attitude and Shape Effects for Non-Resolved Objects," *Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, Maui, HI, 2007, pp. 464–475.

[4] Wetterer, C. J. and Jah, M. K., "Attitude Determination from Light Curves," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 5, Sept.-Oct. 2009, pp. 1648–1651, doi:10.2514/1.44254.

[5] Holzinger, M. J., Alfriend, K. T., Wetterer, C. J., Luu, K. K., Sabol, C., and Hamada, K., "Photometric Attitude Estimation for Agile Space Objects with Shape Uncertainty," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 3, May-June 2014, pp. 921–932, doi:10.2514/1.58002.

[6] Coder, R., Linares, R., and Holzinger, M. J., "Improved Models for Radiometric Attitude Estimation of Agile Space Objects," *AAS/AIAA Space Flight Mechanics Meeting*, AAS, Springfield, VA, Jan.-Feb. 2015, AAS 2015-231.

[7] Hinks, J. C. and Crassidis, J. L., "Angular Velocity Bounds via Light Curve Glint Duration," *AIAA Guidance, Navigation, and Control Conference, AIAA SciTech*, AIAA, Reston, VA, Aug. 2016, AIAA 2016-0627, doi:10.2514/6.2016-0627.

[8] Wetterer, C. J., Chow, C. C., Crassidis, J. L., Linares, R., and Jah, M. K., "Simultaneous Position, Velocity, Attitude, Angular Rates, and Surface Parameter Estimation using Astrometric and Photometric Observations," *2013 16th International Conference on Information Fusion (FUSION)*, IEEE, Piscataway, NJ, 2013, pp. 997–1004.

[9] Kaasalainen, M. and Torppa, J., "Optimization Methods for Asteroid Lightcurve Inversion I: Shape Determination," *Icarus*, Vol. 153, No. 1, Sept. 2001, pp. 24–36, doi:10.1006/icar.2001.6673.

[10] Kaasalainen, M., Torppa, J., and Muinonen, K., "Optimization Methods for Asteroid Lightcurve Inversion II: The Complete Inverse Problem," *Icarus*, Vol. 153, No. 1, Sept. 2001, pp. 37–51, doi:10.1006/icar.2001.6674.

[11] Calef, B., Africano, J., Birge, B., Hall, D., and Kervin, P. W., "Photometric Signature Inversion," *Proceedings of SPIE*, Vol. 6307, Unconventional Imaging II, Aug. 2006, Paper 11, doi: 10.1117/12.683015.

[12] Jah, M. and Madler, R., "Satellite Characterization: Angles and Light Curve Data Fusion for Spacecraft State and Parameter Estimation," *Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, Maui, HI, Sept. 2007, Paper E49.

[13] Metopolis, N., Rosenbluth, A. W., Teller, A. H., and Teller, E., "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, Vol. 21, 1953, pp. 1087–1092, doi:10.1063/1.1699114.

[14] Hastings, W. K., "Monte Carlo Sampling Methods using Markov Chains and Their Applications," *Biometrika*, Vol. 57, No. 1, April 1970, pp. 97–109, doi:10.2307/2334940.

[15] Roberts, G. O. and Rosenthal, J. S., "Examples of Adaptive MCMC," *Journal of Computational and Graphical Statistics*, Vol. 18, No. 2, 2009, pp. 349–367, doi:10.1198/jcgs.2009.06134.

[16]Haario, H., Saksman, E., and Tamminen, J., "An Adaptive Metropolis Algorithm," *Bernoulli*, Vol. 7, No. 2, 2001, pp. 223–242.

[17]Mira, A., "On Metropolis-Hastings Algorithms with Delayed Rejection," *Metron*, Vol. 59, No. 3-4, 2001, pp. 231–241.

[18]Haario, H., Laine, M., Mira, A., and Saksman, E., "DRAM: Efficient Adaptive MCMC," *Statistics and Computing*, Vol. 16, No. 4, Dec. 2006, pp. 339–354, doi:10.1007/s11222-006-9438-0.

[19]Rosenthal, J. and Roberts, G., "Coupling and Ergodicity of Adaptive MCMC," *Journal of Applied Probablity*, Vol. 44, No. 2, June 2007, pp. 458–475, doi:10.2307/27595854.

[20]Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D., "Hybrid Monte Carlo," *Physics Letters B*, Vol. 195, No. 2,3, Sept. 1987, pp. 216–222, doi:10.1016/0370-2693(87)91197-X.

[21]Girolami, M. and Calderhead, B., "Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 73, No. 2, March 2011, pp. 123–214, doi:10.1111/j.1467-9868.2010.00765.x.

[22]Wetterer, C. J., Linares, R., Crassidis, J. L., Kelecy, T. M., Ziebart, M. K., Jah, M. K., and Cefola, P. J., "Refining Space Object Radiation Pressure Modeling with Bidirectional Reflectance Distribution Functions," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 1, 2013, pp. 185–196.

[23]Markley, F. L. and Crassidis, J. L., *Fundamentals of Spacecraft Attitude Determination and Control*, Springer, New York, NY, 2014, p. 345, doi:10.1007/978-1-4939-0802-8.

[24]Crassidis, J. L. and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, CRC Press, Boca Raton, FL, 2nd ed., 2012, pp. 91–92, 103–108.

[25]Ashikmin, M. and Shirley, P., "An Anisotropic Phong Light Reflection Model," Tech. Rep. UUCS-00-014, University of Utah, Salt Lake City, UT, 2000.

[26]Hinks, J. C., Linares, R., and Crassidis, J. L., "Attitude Observability from Light Curve Measurements," *AIAA Guidance, Navigation, and Control Conference*, AIAA, Reston, VA, Aug. 2013, AIAA 2013-5005, doi:10.2514/6.2013-5005.

[27]Tibshirani, R., "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, No. 1, 1996, pp. 267–288, doi:10.2307/2346178.

[28]Mohamed, S., Ghahramani, Z., and Heller, K. A., "Bayesian and L1 Approaches for Sparse Unsupervised Learning," *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, edited by J. Langford and J. Pineau, ACM, New York, NY, 2012, pp. 751–758.

[29]Vrba, F. J., DiVittorio, M. E., Hindsley, R. B., Schmitt, H. R., Armstrong, J. T., Shankland, P. D., Hutter, D. J., and Benson, J. A., "A Survey of Geosynchronous Satellite Glints," *Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, Maui, HI, Sept. 2009, Paper E.28.

[30]Blinn, J. F., "Models of Light Reflection for Computer Synthesized Pictures," *SIGGRAPH '77: Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, 1977, pp. 192–198, doi:10.1145/563858.563893.

[31]Cook, R. L. and Torrance, K. E., "A Reflectance Model for Computer Graphics," *ACM Transactions on Graphics*, Vol. 1, No. 1, Jan. 1982, pp. 7–24, doi:10.1145/357290.357293.