

Toward Guaranteed Illumination Models for Non-Convex Objects

Yuqian Zhang¹, Cun Mu², Han-wen Kuo¹, John Wright¹

¹Department of Electrical Engineering, Columbia University

²Department of Industrial Engineering and Operations Research, Columbia University

August 3, 2018

Abstract

Illumination variation remains a central challenge in object detection and recognition. Existing analyses of illumination variation typically pertain to convex, Lambertian objects, and guarantee quality of approximation in an average case sense. We show that it is possible to build \mathcal{V} (vertex)-description convex cone models with worst-case performance guarantees, for nonconvex Lambertian objects. Namely, a natural verification test based on the angle to the constructed cone guarantees to accept any image which is sufficiently well-approximated by an image of the object under some admissible lighting condition, and guarantees to reject any image that does not have a sufficiently good approximation. The cone models are generated by sampling point illuminations with sufficient density, which follows from a new perturbation bound for point images in the Lambertian model. As the number of point images required for guaranteed verification may be large, we introduce a new formulation for *cone preserving dimensionality reduction*, which leverages tools from sparse and low-rank decomposition to reduce the complexity, while controlling the approximation error with respect to the original cone.

1 Introduction

Illumination variation remains a central challenge in object detection and recognition. Changes in lighting can dramatically change the appearance of the object, rendering simple pattern recognition techniques such as nearest neighbor ineffective. Various approaches have been proposed to mitigate this problem, for example, using nonlinear features based on gradient orientation [Low04], using quotient images [SRR01] or total variation regularization [CZ⁺05]. These approaches are often effective in practice, but can break down under extreme illumination. Moreover, because of the nonlinearity of the feature extraction step, clearly characterizing their domain of applicability is challenging.

An alternative approach is to attempt to explicitly characterize the set of images of the object that can be generated under varying lighting. The seminal work [BK98] argues that images of a given object with fixed pose and varying illumination should lie near a convex cone in the high-dimensional image space. This conic structure arises as a consequence of nonnegativity of light and linearity of light transport. Many subsequent works have attempted to capture the gross structure

of this cone using low-dimensional convex cone or linear subspace models. Motivated by empirical evidence of low-dimensional linear structure in image sets taken under varying illumination (e.g., [EHY95]), [BJ03] and [Ram02] used an elegant interpretation of the Lambertian reflectance as spherical convolution to argue that for a *convex, Lambertian object*, a linear subspace of nine dimensions may suffice to capture most of the variance due to lighting. These models have been used for recognition in many subsequent works [GBK01, WZL⁺, WYG⁺09, WWG⁺12], and have been extended in a number of directions [FSB04, RKB]. The promise of subspace or cone models, compared to feature-based approaches described above, is that, by reasoning carefully about the image formation process, it might be possible to guarantee to well-approximate all images of the object under clearly delineated conditions.

It is worth asking, then, *what approximation guarantees do current results afford us?* For convex, Lambertian objects, it can be shown that for one or more uniformly random point sources, a nine dimensional spherical harmonic approximation captures *on average* about 98% of the energy [BJ03, FSB04]. However, per discussion in [BJ03], low-dimensional linear models do not guarantee quality of approximation for arbitrary extreme illumination conditions. Moreover, for more general nonconvex objects, cast shadows bring in discontinuous changes in radiosity, which render spherical harmonic approximations ineffective [RKB]. Strictly speaking, no rigorous guarantees on quality of approximation are currently known for general nonconvex objects.

In this work, we ask whether it is possible to build models for illumination variation with the following desirable characteristics:

- (i) **UNIFORM GUARANTEES:** Guaranteed robustness to worst case lighting, over some clearly specified class of admissible lighting conditions.
- (ii) **NONCONVEXITY:** Work even for nonconvex objects, with a representation complexity that is adaptive to the complexity of the object of interest.
- (iii) **EFFICIENCY:** Low storage and computational complexity.

We study these questions in the context of a model problem in object instance verification, in which one is given an object \mathcal{O} at a fixed pose, and ask whether the input image is an image of this object under some valid illumination condition. We develop rigorous guarantees for this problem, for general (including nonconvex) Lambertian objects. Our results show how to build a model that guarantees to accept every image that can be interpreted as an image of the object under some lighting condition, and to reject every image that is sufficiently dissimilar to all images of the object under valid lighting conditions.

Similar to [LHK05, MLJ09, WWG⁺12], and many other previous works, we construct a \mathcal{V} -approximation to the illumination cone, which approximates this cone with the conic hull of a finite collection of images taken under point illuminations. Previous empirical work has suggested that the number of images required for an accurate representation can be large [MLJ09]. However, again, for this representation no quantitative results on quality of approximation are currently known. We start from the goal of building a provably correct algorithm for instance verification, and show that in this setting, this reduces to approximating the illumination cone in Hausdorff sense. We derive, in terms of the properties of the object and the scene, sufficient sampling densities for this goal to be met. Our bounds depend on properties of the scene and the object – in particular, they depend on the level of ambient illumination, and a notion of *convexity defect*. They make precise the intuitions that (i) it is more difficult to operate in low-light scenarios, and (ii) nonconvex objects are more challenging than convex objects.

The number of images required to guarantee performance can be large. To address this problem, we introduce a new approach to *cone preserving complexity reduction*. This approach uses tools from convex programming – in particular, sparse and low-rank decomposition [CLMW11, CSPW11] – but introduces a new constrained formulation which guarantees that the conic hull of the output will well-approximate the conic hull of the input. The low-rank and sparse decomposition leverages our *qualitative* understanding of the physical properties of images (low-dimensionality, sparsity of cast shadows) [BJ03, Ram02, WYG⁺09, CLMW11, WGS⁺10], while the constraint ensures that the output of this algorithm is *always* a good approximation to the target cone. Empirically, we find that the output is often of much lower complexity than the input. This suggests a methodology for building instance verifiers that are both robust to worst case illumination, and computationally efficient.

2 Problem Formulation and Methodology

Cone Models for Illumination. We consider images of size $w \times h$, and let $m = wh$. Each image can be treated as a vector $\mathbf{y} \in \mathbb{R}^m$. We are interested in the set of images of an object \mathcal{O} that can be generated under distant illumination. These images form a subset $C_0 \subseteq \mathbb{R}^m$. Each distant illumination can be identified with a nonnegative function $f : \mathbb{S}^2 \rightarrow \mathbb{R}_+$, whose value $f(\mathbf{u})$ is the intensity of light from direction \mathbf{u} . We use the notation \mathcal{F} for the set of nonnegative, Riemann integrable functions on \mathbb{S}^2 .¹ Mathematically, \mathcal{F} is a convex cone: sums of nonnegative, integrable functions are again nonnegative and integrable.

We assume a linear sensor response: the image is a linear function of the incident irradiance.² By linearity of light transport and linearity of the sensor response, the observed image $\mathbf{y} \in \mathbb{R}^m$ is a linear function $\mathbf{y}[f]$ of the illumination f : if the object is subjected to the superposition $f = f_1 + f_2$ of two illuminations f_1 and f_2 , we observe $\mathbf{y}[f_1 + f_2] = \mathbf{y}[f_1] + \mathbf{y}[f_2]$. Since f resides in the convex cone, the set $C_0 \doteq \mathbf{y}[\mathcal{F}] \subset \mathbb{R}^m$ of possible images is also a convex cone. Note, however, that the fact that C_0 is a convex cone holds under very mild assumptions.

The detailed properties of C_0 were first studied in [BK98], and a great deal of subsequent work has been devoted to understanding its properties [Ram02, BJ03, FSB04]. Most of this body of work has been devoted to simple, analytically tractable models such as convex, Lambertian objects. As discussed above, for such simple models, interesting qualitative statements can be made about the gross shape of C_0 .

The cone C_0 can be interpreted as the set of all images of the object under different distant lighting conditions. Intuitively speaking, we expect the problem of representing images \mathbf{y} under different illuminations to be more challenging when the light has a stronger directional component. To capture the relative contribution of directional and ambient components of light, we introduce a family of function classes \mathcal{F}_α , indexed by parameter $\alpha \in [0, \infty)$. Illuminations in \mathcal{F}_α consist of an ambient component $\alpha\omega$, where $\omega(\mathbf{u}) = 1/\text{area}(\mathbb{S}^2)$ is the constant function on the sphere, and an

¹To be clear, we call f Riemann integrable iff it is integrable in spherical coordinates: writing $W = [0, 2\pi] \times [0, \pi]$ and $\eta : W \rightarrow \mathbb{S}^2$ via $\eta(\theta, \phi) = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi)$, f is Riemann integrable iff $f \circ \eta \sin \phi$ is Riemann integrable as a function on $W \subseteq \mathbb{R}^2$. We let $\int_{\mathbf{u}} f(\mathbf{u}) d\mathbf{u} = \int_W f \circ \eta \sin \phi d(\theta, \phi)$, where the right hand side is a Riemann integral. We reserve the related notation $\int f(\mathbf{u}) d\sigma(\mathbf{u})$ for the (Lebesgue) integral with respect to the spherical measure. When $f \in \mathcal{F}$, these two integrals coincide.

²This model neglects saturation and quantization.

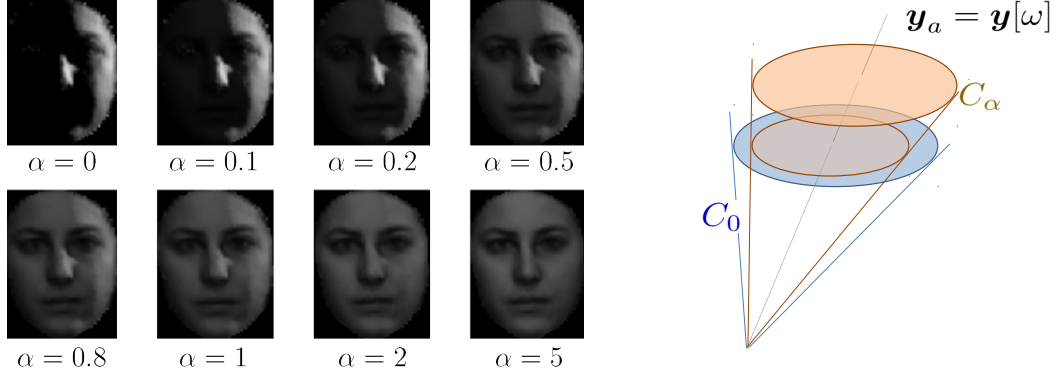


Figure 1: **Ambient level α .** Left: typical images from the cone C_α , for ambient levels $\alpha = 0$ up to $\alpha = 5$. In each example f_d is an extreme directional illumination. Images rendered from [SSTB12]. Right: illumination cones C_α with varying ambient level α .

arbitrary (possibly directional) component f_d :

$$\mathcal{F}_\alpha = \{f_d + \alpha\omega \mid f_d \in \mathcal{F}, \|f_d\|_{L_1} \leq 1\}, \quad (2.1)$$

For each ambient level α , we have a cone

$$C_\alpha \doteq \mathbb{R}_+ \cdot \mathbf{y}[\mathcal{F}_\alpha]. \quad (2.2)$$

For any $\alpha \leq \alpha'$, $C_{\alpha'} \subseteq C_\alpha$. In this sense, the choice of α induces a tradeoff: as α becomes smaller, C_α becomes more complicated to compute with, but can represent broader illumination conditions. Our complexity bounds in Section 5 will make this intuition precise. Figure 1 shows rendered images of a face under various ambient levels $\alpha \geq 0$. Our methodology is compatible with any choice of $\alpha > 0$.

Verification using Convex Cones. Our methodology asks the system designer to select a target level of ambient illumination α , and hence choose a target cone $C = C_\alpha$. At test time, we are given a new input image $\mathbf{y} \in \mathbb{R}^m$. The verification problem asks us to decide if \mathbf{y} could be an image of object \mathcal{O} : *Is \mathbf{y} an element of C ?* Real images contain noise and other imperfections. Hence, in practice, it is more appropriate to ask whether \mathbf{y} is *sufficiently close* to C . The distance from \mathbf{y} to C in ℓ^2 -norm is

$$d(\mathbf{y}, C) \doteq \inf \{\|\mathbf{y} - \mathbf{y}'\|_2 \mid \mathbf{y}' \in C\}.$$

Any cone C is nonnegatively homogeneous: if $\mathbf{z} \in C$, $t\mathbf{z} \in C$ for all $t \geq 0$. To obtain a criterion for verification which is scale invariant, rather than working directly with the distance $d(\mathbf{y}, C)$, we work with the angle

$$\angle(\mathbf{y}, C) \doteq \text{asin} \left(\frac{d(\mathbf{y}, C)}{\|\mathbf{y}\|_2} \right).$$

This leads to a simple, natural criterion for verification:

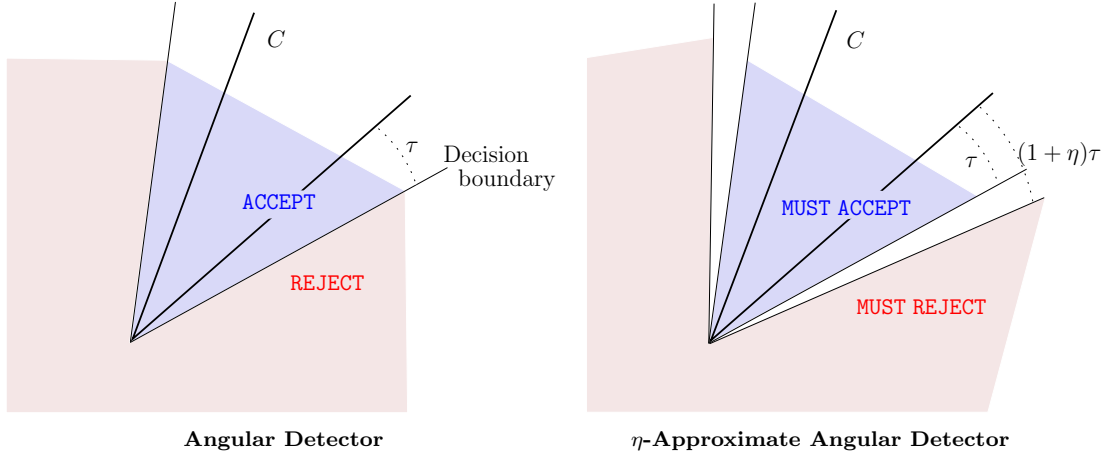


Figure 2: **Two detection rules.** The angular detector accepts points based on their angle with the cone C . An approximate angular detector guarantees to accept any point within angle τ of C , and to reject any point with angle greater than $(1 + \eta)\tau$. In the intermediate region (white) there are no restrictions on its behavior.

Definition 2.1. *The angular detector (AD) $\mathfrak{D}_\tau^C : \mathbb{R}_+^m \rightarrow \{\text{ACCEPT}, \text{REJECT}\}$ with threshold τ is the decision rule*

$$\mathfrak{D}_\tau^C(\mathbf{y}) = \begin{cases} \text{ACCEPT} & \angle(\mathbf{y}, C) \leq \tau, \\ \text{REJECT} & \angle(\mathbf{y}, C) > \tau. \end{cases} \quad (2.3)$$

This rule has a simple interpretation: we accept \mathbf{y} if and only if it can be interpreted as an image of \mathcal{O} plus a noise perturbation, and the signal-to-noise ratio is sufficiently large.

If C is a polyhedral cone, the decision rule (2.3) can be implemented via nonnegative least squares. This is efficient if the number n of extreme rays of C is small. If \mathcal{O} is a convex polyhedron with only a few faces, this is the case. However, in general, the number of extreme rays in a \mathcal{V} (vertex)-description can be large or even unbounded.³ One remedy is to relax the definition slightly:

Definition 2.2. *$f : \mathbb{R}_+^m \rightarrow \{\text{ACCEPT}, \text{REJECT}\}$ is an η -approximate angular detector (η -AAD) if*

$$f(\mathbf{y}) = \begin{cases} \text{ACCEPT} & \angle(\mathbf{y}, C) \leq \tau, \\ \text{REJECT} & \angle(\mathbf{y}, C) > (1 + \eta)\tau. \end{cases} \quad (2.4)$$

We let $\hat{\mathbb{D}}_{\tau, \eta}^C$ denote the set of all such f .

Figure 2 displays the AD and its η -relaxation. We can regard η -AAD as a relaxed version of AD in the sense that when $\angle(\mathbf{y}, C) \in (\tau, (1 + \eta)\tau]$, no demands are placed on the output of the algorithm. This buffer zone allows us to work with a surrogate cone \hat{C} with much simpler structure, enabling computationally tractable (even efficient!) verification. For example, if we

³For convex, Lambertian objects, in a point sampling model of image formation, the best known bound on the number of extreme rays in a \mathcal{V} -representation of C is quadratic in the number of image pixels: $n = O(m^2)$ [BK98]. For nonconvex objects or more realistic sampling models, C may not even be polyhedral.

form a polyhedral approximation $\widehat{C} = \text{cone}(\widehat{\mathbf{A}})$, the distance to \widehat{C} is just the optimal value of the nonnegative least squares problem

$$d(\mathbf{y}, \widehat{C}) = \min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - \widehat{\mathbf{A}}\mathbf{x}\|_2^2. \quad (2.5)$$

To implement the angular detector $\mathfrak{D}_\xi^{\widehat{C}}$ for \widehat{C} , we just need to solve (2.5) and compare the optimal value to a threshold.

It should come as no surprise that whenever \widehat{C} approximates C sufficiently well, we have detector $\mathfrak{D}_\xi^{\widehat{C}} \in \widehat{\mathbb{D}}_{\tau, \eta}^C$, with ξ chosen appropriately. In words, applying the angular test with \widehat{C} gives an approximate angular detector for the original cone C . To make this precise, we need a notion of approximation. We will work with the following discrepancy δ :

$$\delta(C, \widehat{C}) = \max \left\{ \sup_{\mathbf{y} \in C, \|\mathbf{y}\|=1} d(\mathbf{y}, \widehat{C}), \sup_{\mathbf{y} \in \widehat{C}, \|\mathbf{y}\|=1} d(\mathbf{y}, C) \right\}. \quad (2.6)$$

This is just the Hausdorff distance between $C \cap \mathfrak{B}(\mathbf{0}, 1)$ and $\widehat{C} \cap \mathfrak{B}(\mathbf{0}, 1)$. It therefore satisfies the triangle inequality: $\forall \bar{C}$,

$$\delta(C, \widehat{C}) \leq \delta(C, \bar{C}) + \delta(\bar{C}, \widehat{C}). \quad (2.7)$$

If $\delta(C, \widehat{C})$ is small, we indeed lose little in working with \widehat{C} :

Lemma 2.3. *Given cone C , $\tau > 0$ and $\eta \geq 0$ with $(1 + \eta)\tau \in (0, \frac{\pi}{2})$, and another cone \widehat{C} , we have $\mathfrak{D}_\xi^{\widehat{C}} \in \widehat{\mathbb{D}}_{\tau, \eta}^C$ whenever*

$$\delta(C, \widehat{C}) \leq \frac{1}{2} (\sin(\tau + \eta\tau) - \sin \tau) \quad (2.8)$$

and

$$\xi \in \left[\text{asin} \left(\sin \tau + \delta(C, \widehat{C}) \right), \text{asin} \left(\sin(\tau + \eta\tau) - \delta(C, \widehat{C}) \right) \right]. \quad (2.9)$$

Proof. Please see Appendix A. □

So, if $\delta(C, \widehat{C})$ is small, we can simply apply an angular test with cone \widehat{C} , and this will implement an approximate angular detector for C . Notice that whenever (2.8) is satisfied, we may satisfy (2.9) by setting $\xi = \text{asin} \left(\frac{1}{2} \sin \tau + \frac{1}{2} \sin(\tau + \eta\tau) \right)$.

Goals and Methodology. From the above discussion, if we want to provide a detector that guarantees to accept any image that has a valid interpretation as an image of the object under some lighting $f \in \mathcal{F}_\alpha$, and reject any image that cannot be plausibly interpreted as an image under $f \in \mathcal{F}_\alpha$, it is enough to build an approximation \widehat{C} to the cone C_α , and the correct notion of approximation is the Hausdorff distance. The question, then, is how to build such an approximation: how complicated does \widehat{C} have to be to guarantee $\delta(\widehat{C}, C_\alpha) \leq \gamma$? This leads to a way of formalizing several fundamental questions in illumination-robust detection and recognition: *What information do we need to guarantee robust verification performance? How does this sample complexity depend on the complexity of the class of illuminations the system must handle? How does it depend on the properties of the object?*

In the sequel, we will show how to build a \mathcal{V} -approximation $\bar{C} = \text{cone}(\bar{\mathbf{A}})$ to C_α , where $\bar{\mathbf{A}} \in \mathbb{R}^{m \times n}$ is a matrix whose columns are images under point illumination. The underlying question is how

large n needs to be, in terms of ambient illumination level α and the desired quality of approximation ε . We will show that for Lambertian objects,

$$n(\alpha, \varepsilon) = \frac{\text{const}(\text{sensor}, \text{object})}{\alpha^4 \varepsilon^4} \quad (2.10)$$

examples suffice. The numerator depends only on physical properties of the object and of the imaging system, which we will make precise below. It is worth remarking that the fact that a polynomial dependence on ε^{-1} is possible at all can be considered remarkable here – this is certainly not the case for general high-dimensional convex cones. The reason that such a result is possible at all is that the extreme rays of our cone of interest will turn out to have much lower dimensional structure: they are generated by point illuminations, which are indexed by the sphere. Turning this intuition into a rigorous result will require a detailed analysis of the extreme rays of C , which we carry out below.

Section 3 characterizes the extreme rays of the cone C_α . Section 4 describes our imaging model in detail. Section 5 describes several new perturbation bounds which lead to the estimate of sample complexity (2.10). In Section 6, by solving a convex optimization problem, we form cone \hat{C} , a γ -approximation to \bar{C} , but with much lower complexity. From (2.7), our resulting cone \hat{C} ($\varepsilon + \gamma$)-approximates C_α : $\delta(C_\alpha, \hat{C}) \leq \varepsilon + \gamma$. Finally, Section 7 presents several numerical experiments.

3 Extreme Rays of C_α

In the previous section, we saw that for guaranteed verification with a cone C , it was enough to approximate that cone in Hausdorff sense. For computational purposes, perhaps the most natural approximation is a \mathcal{V} (vertex) approximation – we would like to write $\bar{C} = \text{cone}(\bar{\mathbf{A}})$ for some matrix $\bar{\mathbf{A}}$. To this end, we need to characterize the extreme rays of C_α , for $\alpha \geq 0$. We will see below⁴ that for our models of interest, the linear function $\mathbf{y}[f]$ can be written as

$$\mathbf{y}[f] = \int_{\mathbf{u} \in \mathbb{S}^2} \bar{\mathbf{y}}[\mathbf{u}] f(\mathbf{u}) d\mathbf{u}, \quad (3.1)$$

where $\bar{\mathbf{y}} : \mathbb{S}^2 \rightarrow \mathbb{R}^m$ is a continuous function. In this expression, we have used the natural extension of the Riemann integral to vector-valued functions on the sphere, which simply integrates each of the m coordinate functions.

We begin by characterizing the extreme rays of $C_0 = \mathbf{y}[\mathcal{F}]$. These turn out to simply be the vectors $\bar{\mathbf{y}}[\mathbf{u}]$:

Lemma 3.1. *Suppose that the imaging map \mathbf{y} satisfies (3.1), with $\bar{\mathbf{y}}[\cdot] : \mathbb{S}^2 \rightarrow \mathbb{R}^m$ continuous. Then if $C_0 = \mathbf{y}[\mathcal{F}]$,*

$$\delta(C_0, \text{cone}(\{\bar{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\})) = 0. \quad (3.2)$$

Proof. Please see Appendix B. □

In the physical imaging models we consider, the $\bar{\mathbf{y}}[\mathbf{u}]$ can be considered images of \mathcal{O} under *point illumination* from direction \mathbf{u} . With this interpretation, the previous lemma simply asserts that any image $\mathbf{y}[f]$ under distant, Riemann integrable illumination f can be arbitrarily well approximated

⁴For a rigorous argument, please see section 5.

using a conic combination of images under point illumination.⁵ The conic hull of these extreme images is equal to the cone C_0 of images of \mathcal{O} under arbitrary Riemann integrable illumination, up to a set of measure zero.

We would like a similar expression that works when the ambient level is larger than zero – we would like to also approximate the extreme rays of C_α . The following lemma says that that we can use images of the form $\check{\mathbf{y}}[\mathbf{u}] = \bar{\mathbf{y}}[\mathbf{u}] + \alpha \mathbf{y}_a$, where \mathbf{y}_a is the image of \mathcal{O} under ambient illumination:

Lemma 3.2. *Suppose that $\mathbf{y}[f]$ satisfies (3.1) with $\bar{\mathbf{y}}[\cdot]$ continuous. Set $\check{\mathbf{y}}[\mathbf{u}] = \bar{\mathbf{y}}[\mathbf{u}] + \alpha \mathbf{y}_a$, with*

$$\mathbf{y}_a = \frac{1}{\text{area}(\mathbb{S}^2)} \int_{\mathbf{u}} \bar{\mathbf{y}}[\mathbf{u}] d\mathbf{u}, \quad (3.3)$$

and $\check{C} = \text{cone}(\{\check{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\})$. Then, we have $\delta(C_\alpha, \check{C}) = 0$.

Proof. Please see Appendix B. □

This lemma says that to work with C_α , we can simply work with a modified set of extreme images $\check{\mathbf{y}}[\mathbf{u}]$, which are sums of images under point illumination and the ambient image \mathbf{y}_a . We still need to build a computationally tractable representation for C_α . A natural approach is to discretize the set of illumination directions, by choosing a finite set $\mathbf{u}_1, \dots, \mathbf{u}_N$. The following lemma asserts that as long as the $\bar{\mathbf{y}}[\mathbf{u}_i]$ can approximate any point illumination $\bar{\mathbf{y}}[\mathbf{u}]$ in an absolute sense, the cone generated by the finite set and the cone C_α will not differ too much:

Lemma 3.3. *Let $\bar{C} = \text{cone}(\check{\mathbf{y}}[\mathbf{u}_1], \dots, \check{\mathbf{y}}[\mathbf{u}_N], \mathbf{y}_a)$, and*

$$\delta(C_\alpha, \bar{C}) = \delta(\check{C}, \bar{C}) \leq \frac{2 \sup_{\mathbf{u} \in \mathbb{S}^2} \min_i \|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}_i]\|_2}{\eta_\star \alpha \|\mathbf{y}_a\|_2}. \quad (3.4)$$

here $\eta_\star = \sup_{\|\mathbf{w}\|_2 \leq 1} \inf_{\mathbf{u}} \left\langle \mathbf{w}, \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} \right\rangle \geq m^{-1/2}$ measures the angular spread of C_α .

Proof. Please see Appendix B. □

This substantially simplifies the problem of approximating C_α : to control the error over *all* possible images, it is enough to control the error over images under *point illumination*. Below, we will see that this is possible, even for nonconvex objects, provided the object's reflectance is Lambertian.

4 Physical Assumptions: Lambertian Objects

We will introduce a set of hypotheses on the object and the image formation process. Under these hypotheses, we obtain rigorous bounds for the error $\|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2$ incurred by approximating an image $\bar{\mathbf{y}}[\mathbf{u}]$ under point illumination \mathbf{u} with another image $\bar{\mathbf{y}}[\mathbf{u}']$ under point illumination \mathbf{u}' . From the results in the previous sections, a good approximation of images under distant point illumination will be sufficient to ensure a good approximation to the cone of all images of the object under distant illumination in Hausdorff sense.

⁵Informal variants of Lemma 3.1 are stated in many previous works in this area; see, e.g., [BK98].

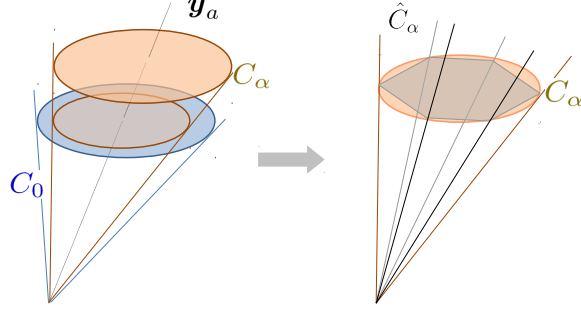


Figure 3: **Cone Approximation:** C_α (left) and its \mathcal{V} -approximation (right).

Object Geometry. Our bounds pertain to *triangulated* objects, whose boundary is a union of finitely many oriented triangles:

Definition 4.1 (Triangulated object). *We say that $\mathcal{O} \subset \mathbb{R}^3$ is triangulated if for some integer N ,*

$$\begin{aligned} \partial\mathcal{O} &= \bigcup_{i=1}^N \Delta_i, \\ \forall i, \Delta_i &= \text{conv} \left\{ \mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)}, \mathbf{v}_i^{(3)} \right\}, \dim(\Delta_i) = 2, \\ \forall i \neq j, \Delta_i \cap \Delta_j &\in \{\emptyset\} \cup \mathcal{V} \cup \mathcal{E}, \end{aligned}$$

where \mathcal{V} and \mathcal{E} are sets of vertices and edges:

$$\mathcal{V} = \left\{ \left\{ \mathbf{v}_i^{(k)} \right\} \mid i \in [N], k \in [3] \right\}, \quad \mathcal{E} = \left\{ \text{conv} \left\{ \mathbf{v}_i^{(k_1)}, \mathbf{v}_i^{(k_2)} \right\} \mid i \in [N], k_1 \neq k_2 \right\},$$

and each face Δ_i has a unique outward normal $\mathbf{n}_i \in \mathbb{S}^2$.

This geometric assumption captures most of the object models that are interesting for computer graphics and vision. Notice that N above can be arbitrarily large – and hence this model can approximate smooth objects.

The normal vectors \mathbf{n}_i play an important role in describing how light interacts with the object. If we let

$$\Phi = \bigcup_{\Delta_i} \text{relint}(\Delta_i) \tag{4.1}$$

be the union of the relative interiors of faces of the object, for $\mathbf{x} \in \Phi$, the outward normal \mathbf{n} is uniquely defined, and we can write it as $\mathbf{n}(\mathbf{x}) \in \mathbb{S}^2$. We write

$$E \doteq \partial\mathcal{O} \setminus \Phi = \bigcup_{e \in \mathcal{E}} e \tag{4.2}$$

for the remaining points. This is the set of all points contained in some edge e .

We will introduce two indicator functions that describe how object obstructs the “view” from a given point $\mathbf{x} \in \partial\mathcal{O}$:

- The *point-direction visibility indicator* $\nu : \partial\mathcal{O} \times \mathbb{S}^2 \rightarrow \{0, 1\}$ indicates those directions \mathbf{u} , which when viewed from point \mathbf{x} , are not obstructed by other points of the object:

$$\nu(\mathbf{x}, \mathbf{u}) = \begin{cases} 1 & (\{\mathbf{x}\} + \mathbb{R}_+\mathbf{u}) \cap \mathcal{O} = \{\mathbf{x}\}, \\ 0 & \text{else.} \end{cases} \quad (4.3)$$

- The *point-point visibility indicator* $V : \partial\mathcal{O} \times \partial\mathcal{O} \rightarrow \{0, 1\}$ indicates those point pairs $(\mathbf{x}, \mathbf{x}') \in \partial\mathcal{O} \times \partial\mathcal{O}$ that are mutually visible:

$$V(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & [\mathbf{x}, \mathbf{x}'] \cap \mathcal{O} = \{\mathbf{x}, \mathbf{x}'\}, \\ 0 & \text{else.} \end{cases} \quad (4.4)$$

Integrating on $\partial\mathcal{O}$. To clearly describe how light interacts with the object \mathcal{O} to produce an image, we need to be able to integrate on $\partial\mathcal{O}$. This is conceptually straightforward. In this section, we simply introduce notation for this integral; a detailed construction is given in Appendix C. There, we formally construct a measure space $(\Phi, \Sigma_{\partial\mathcal{O}}, \mu_{\partial\mathcal{O}})$. For $g : \partial\mathcal{O} \rightarrow \mathbb{R}$, the Lebesgue integral with respect to this measure will be written as

$$\int g(\mathbf{x}) d\mu_{\partial\mathcal{O}}(\mathbf{x}). \quad (4.5)$$

We can define a vector space

$$L^2[\partial\mathcal{O}] = \{g : \partial\mathcal{O} \rightarrow \mathbb{R} \mid g^2 \text{ is integrable} \}. \quad (4.6)$$

For $g \in L^2[\partial\mathcal{O}]$, we define

$$\|g\|_{L^2} \doteq \left(\int g(\mathbf{x})^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) \right)^{1/2}. \quad (4.7)$$

Object Reflectance. We will consider a Lambertian reflectance model. In this model, the object is fully described by its geometry and its *albedo*

$$\rho : \partial\mathcal{O} \rightarrow (0, 1], \quad (4.8)$$

which is the fraction of incoming light that is reflected at each point $\mathbf{x} \in \partial\mathcal{O}$. We assume that the albedo is positive everywhere, and that it is $\Sigma_{\partial\mathcal{O}}$ -measurable. In the Lambertian model, the key quantity linking the illumination f and the image \mathbf{y} is the outgoing irradiance (radiosity) at each point $\mathbf{x} \in \partial\mathcal{O}$:

$$g : \partial\mathcal{O} \rightarrow \mathbb{R}_+. \quad (4.9)$$

Informally speaking, the irradiance $g(\cdot)$ is generated as follows: light from the source impinges on the surface of the object; some is absorbed, while some is reflected. This reflected light can itself illuminate the object, as can further reflections of the reflected light. Then the map from distant illumination f to outgoing irradiance g can be described in terms of two operators.

The **direct illumination** operator $\mathcal{D} : L^2[\mathbb{S}^2] \rightarrow L^2[\partial\mathcal{O}]$ describes the object's reflectance after the first bounce of light from illumination function $f(\mathbf{u})$:

$$\mathcal{D}[f](\mathbf{x}) = \int \bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}) f(\mathbf{u}) d\sigma(\mathbf{u}), \quad \mathbf{x} \in \Phi. \quad (4.10)$$

Here, $\sigma(\cdot)$ is the spherical measure. For Lambertian objects, direct reflectance under point illumination $\bar{\mathcal{D}}$ can be expressed as:

$$\bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}) = \begin{cases} \rho(\mathbf{x}) \langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle_+ \nu(\mathbf{x}, \mathbf{u}) & \mathbf{x} \in \Phi, \\ 0 & \text{else.} \end{cases} \quad (4.11)$$

The **interreflection** operator $\mathcal{T}: L^2[\partial\mathcal{O}] \rightarrow L^2[\partial\mathcal{O}]$ describes how light reflected off the object illuminates the object itself again:

$$\mathcal{T}[g](\mathbf{x}) = \begin{cases} \int \kappa(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') d\mu_{\partial\mathcal{O}}(\mathbf{x}') & \mathbf{x} \in \Phi \\ 0 & \mathbf{x} \in E = \partial\mathcal{O} \setminus \Phi, \end{cases} \quad (4.12)$$

where the kernel κ is given by

$$\kappa(\mathbf{x}, \mathbf{x}') = \frac{\rho(\mathbf{x})}{\pi} \frac{\langle \mathbf{n}(\mathbf{x}'), \mathbf{x} - \mathbf{x}' \rangle \langle \mathbf{n}(\mathbf{x}), \mathbf{x}' - \mathbf{x} \rangle}{\|\mathbf{x} - \mathbf{x}'\|^4} V(\mathbf{x}, \mathbf{x}'). \quad (4.13)$$

For all of the models that we consider, the operator norm of \mathcal{T} will be strictly smaller than one, and so the operator $\mathcal{I} - \mathcal{T}$ will be invertible. Under this assumption the outgoing irradiance on the surface of the object can be written as a convergent series

$$\begin{aligned} g[f] &= \mathcal{D}[f] + \mathcal{T}\mathcal{D}[f] + \mathcal{T}^2\mathcal{D}[f] + \dots \\ &= (\mathcal{I} - \mathcal{T})^{-1}\mathcal{D}[f]. \end{aligned} \quad (4.14)$$

Sensor Model. We consider a perspective camera, with a thin lens model commonly adopted in computer vision with focal length f_c and lens diameter d_c [Hor86].⁶ We assume the imaging sensor is composed of m non-overlapping squares I_i with side length s_c , then the value of the i -th pixel is generated by integrating the irradiance over region I_i :

$$y_i = \mathcal{P}_i[g] \doteq \frac{\gamma_c}{4} \left(\frac{d_c}{f_c} \right)^2 \int_{\mathbf{z} \in I_i} g(\mathbf{p}^{-1}(\mathbf{z})) \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|_2}, \mathbf{e}_3 \right\rangle^4 d\mu(\mathbf{z}). \quad (4.15)$$

Here, \mathbf{p} represents perspective projection; its inverse maps an image point to the corresponding point on $\partial\mathcal{O}$ and γ_c is the camera gain. Combining the expressions for pixels $1 \dots m$, we can describe the image vector as a whole as a linear function of g via

$$\mathbf{y} = \mathcal{P}[g] = \begin{bmatrix} \mathcal{P}_1[g] \\ \vdots \\ \mathcal{P}_m[g] \end{bmatrix} \in \mathbb{R}^m. \quad (4.16)$$

⁶The main idealization in the model (4.15) is that it neglects defocus due to depth. In fact, our methodology is compatible with more sophisticated imaging models, as well as simpler idealizations such as orthographic models. However, the bounds claimed in Lemma 5.5 will change.

The Imaging Operator. Combining the definitions and descriptions in the previous paragraphs, we can give a description of the imaging operator $\mathbf{y}[f]$ as whole. When $\|\mathcal{T}\| < 1$ (i.e., the object is not perfectly reflective), we have

$$\mathbf{y}[f] = \mathcal{P} \sum_{i=0}^{\infty} \mathcal{T}^i \mathcal{D}(f) = \mathcal{P}(I - \mathcal{T})^{-1} \mathcal{D}[f]. \quad (4.17)$$

Using the definition of $\mathcal{D}[\cdot]$ and $\bar{\mathcal{D}}[\cdot]$, we have

Lemma 4.2. *Under the imaging model (4.17), with \mathcal{P} as in (4.16), \mathcal{T} as in (4.12) and \mathcal{D} as in (4.11), if $\|\mathcal{T}\|_{L^2 \rightarrow L^2} < 1$, then for any Riemann integrable f we have*

$$\mathbf{y}[f] = \int \bar{\mathbf{y}}[\mathbf{u}] f(\mathbf{u}) d\mathbf{u}, \quad (4.18)$$

with

$$\bar{\mathbf{y}}[\mathbf{u}] = \mathcal{P}(\mathcal{I} - \mathcal{T})^{-1} \bar{\mathcal{D}}[\mathbf{u}]. \quad (4.19)$$

The quantity $\bar{\mathbf{y}}[\mathbf{u}] \in \mathbb{R}^m$ in Lemma 4.2 can be interpreted as the image of \mathcal{O} under point illumination from direction \mathbf{u} . We will see below that under reasonable hypotheses, $\bar{\mathbf{y}}[\mathbf{u}]$ is continuous in \mathbf{u} . From Lemma 3.3, if we can approximate these $\bar{\mathbf{y}}[\mathbf{u}]$ well, we will well-approximate the cone as a whole.

This proof of the lemma uses Fubini's theorem and monotone convergence to change the order of integration, and then uses the fact that $\bar{\mathbf{y}}[\mathbf{u}]$ is continuous in \mathbf{u} to conclude that the integrand in (4.18) is Riemann integrable. The continuity of $\bar{\mathbf{y}}[\cdot]$ will follow from perturbation bounds in the next section. In Appendix D, we use these results to give a formal proof of Lemma 4.2.

5 Perturbation Bounds and Sufficient Sample Densities

Based on the assumptions laid out above, we will discuss the properties of the linear operators \mathcal{P} , \mathcal{T} , and \mathcal{D} , and show how to control $\|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2$ in terms of $\|\mathbf{u} - \mathbf{u}'\|_2$. The relationship between $\bar{\mathbf{y}}[\mathbf{u}]$ and \mathbf{u} obviously depends on detailed properties of the object \mathcal{O} . In particular, it depends on two complementary quantities measuring the *nonconvexity* of \mathcal{O} :

The **pointwise visibility** is fraction of directions that are visible at point \mathbf{x} , weighted by $\langle \mathbf{n}, \mathbf{u} \rangle$:

$$\tilde{\nu}(\mathbf{x}) \doteq \frac{1}{\pi} \int_{\langle \mathbf{u}, \mathbf{n}(\mathbf{x}) \rangle \geq 0} \langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle \nu(\mathbf{x}, \mathbf{u}) d\sigma(\mathbf{u}) \in [0, 1]. \quad (5.1)$$

here ν is the point-direction visibility indicator function in equation (4.3).

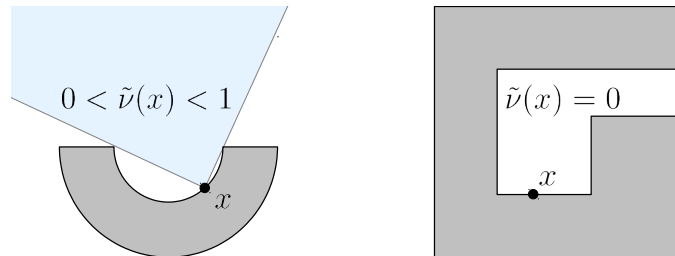


Figure 4: **Pointwise Visibility** $\tilde{\nu}(\mathbf{x})$

The pointwise visibility $\nu(\mathbf{x})$ is a localized nonconvexity measurement, depending on properties of the object perceived from a point \mathbf{x} : smaller value of this quantity suggests more complex geometry around \mathbf{x} . For convex objects, $\nu(\mathbf{x}) = 1$ for any point $\mathbf{x} \in \partial\mathcal{O}$.

The other crucial quantity is the total length of the edges that cast shadows on \mathcal{O} itself, when \mathcal{O} is illuminated from direction \mathbf{u} . We call this the **gnomon length** associated with direction \mathbf{u} .⁷ We reserve the notation $\chi[\mathbf{u}]$ for the collection of edges that cast shadows, when the object is illuminated from direction \mathbf{u} . We will define this quantity formally in the next section, after we have introduced some necessary technical machinery. For now, Figure 5(left) gives a visual example of $\chi[\mathbf{u}]$: the edges in this set are highlighted in yellow.

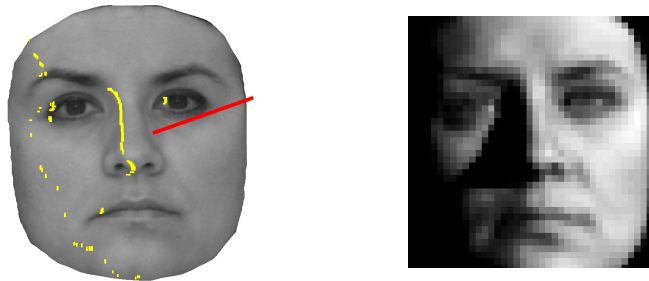


Figure 5: **Shadowing Edges** $\chi[\mathbf{u}]$ (yellow) under point illumination \mathbf{u} (red), with corresponding image on the right.

Compared to $\nu(\mathbf{x})$, $\chi[\mathbf{u}]$ is a more global measurement of nonconvexity, depending on the overall geometry of the object: longer *shadowing edges length* implies more apparent cast shadows. For convex objects, $\chi[\mathbf{u}] = 0$ always holds for any illumination direction $\mathbf{u} \in \mathbb{S}^2$. For nonconvex objects, this quantity helps to bound the change in the image induced by cast shadows, which is a source of considerable difficulty. To state our results more precisely, we begin by introducing some notations and technical machinery for reasoning about the boundary of the shadow region.

5.1 Shadow Boundaries

Under lighting direction \mathbf{u} , the region that is shadowed (not illuminated) is⁸

$$S[\mathbf{u}] \doteq \text{supp}(\bar{\mathcal{D}}[\mathbf{u}])^c.$$

We would like to talk about the boundary of the shadowed region. The follow lemma, which says that the shadowed region $S[\mathbf{u}]$ is closed in the relative topology on $\partial\mathcal{O}$, allows us to do so:

Lemma 5.1. *For all $\mathbf{u} \in \mathbb{S}^2$, $S[\mathbf{u}] \subseteq \partial\mathcal{O}$ is a relatively closed set.*

Proof. Please see Appendix E. □

⁷The “gnomon” is the part of a sundial that casts the shadow.

⁸Here, the *support* $\text{supp}(f) = \{\mathbf{x} \mid f(\mathbf{x}) \neq 0\}$ of a function is its set of nonzeros.

With this in mind, we can let

$$\partial S[\mathbf{u}] \doteq \text{relbdy}(S[\mathbf{u}]) \quad (5.2)$$

denote the shadow boundary, and note that $\partial S[\mathbf{u}] \subseteq S[\mathbf{u}]$. Points on the shadow boundary $\partial S[\mathbf{u}]$ can be separated into those that come from cast shadows and those that come from attached shadows. For $\mathbf{x} \in \partial \mathcal{O}$ and $\mathbf{u} \in \mathbb{S}^2$, let

$$t_\star(\mathbf{x}, \mathbf{u}) \doteq \inf \{t > 0 \mid \mathbf{x} - t\mathbf{u} \in \partial \mathcal{O}\} \in [0, +\infty], \quad (5.3)$$

$$t^\star(\mathbf{x}, \mathbf{u}) \doteq \inf \{t > 0 \mid \mathbf{x} + t\mathbf{u} \in \partial \mathcal{O}\} \in [0, +\infty], \quad (5.4)$$

where we adopt the standard convention that the infimum of the empty set is $+\infty$. We set

$$\mathbf{x}_\mathbf{u} = \mathbf{x} - t_\star(\mathbf{x}, \mathbf{u})\mathbf{u}, \quad \forall (\mathbf{x}, \mathbf{u}) \text{ s.t. } t_\star(\mathbf{x}, \mathbf{u}) < +\infty. \quad (5.5)$$

$$\mathbf{x}^\mathbf{u} = \mathbf{x} + t^\star(\mathbf{x}, \mathbf{u})\mathbf{u}, \quad \forall (\mathbf{x}, \mathbf{u}) \text{ s.t. } t^\star(\mathbf{x}, \mathbf{u}) < +\infty. \quad (5.6)$$

We call $\mathbf{x}_\mathbf{u}$ the *shadow projection* of \mathbf{x} along direction \mathbf{u} , and $\mathbf{x}^\mathbf{u}$ the *shadow retraction* of \mathbf{x} along direction \mathbf{u} . For light direction \mathbf{u} , the physical interpretation of the shadow projection of \mathbf{x} is that it is the first point that is shadowed by \mathbf{x} . Conversely, the shadow retraction is the first point that could cast a shadow on \mathbf{x} . In particular,

$$\nu(\mathbf{x}, \mathbf{u}) = 0 \iff t^\star(\mathbf{x}, \mathbf{u}) < +\infty. \quad (5.7)$$

Notice that because \mathcal{O} is closed, whenever they exist, we have $\mathbf{x}_\mathbf{u} \in \partial \mathcal{O}$ and $\mathbf{x}^\mathbf{u} \in \partial \mathcal{O}$.

The notion of a shadow retraction allows us to associate to each point \mathbf{x} that lies in a cast shadow a point $\mathbf{x}^\mathbf{u}$ which prevents the source from directly illuminating \mathbf{x} . In particular, if \mathbf{x} is in the boundary of a cast shadow, we will see that $\mathbf{x}^\mathbf{u}$ is necessarily an edge point: $\mathbf{x}^\mathbf{u} \in E$. The following technical lemma carries this through precisely:

Lemma 5.2. *Set $C[\mathbf{u}] \doteq \partial S[\mathbf{u}] \cap \Phi$. Then for each $\mathbf{x} \in C[\mathbf{u}]$, $\mathbf{x}^\mathbf{u}$ exists. If we let $\chi[\mathbf{u}] \doteq \{\mathbf{x}^\mathbf{u} \mid \mathbf{x} \in C[\mathbf{u}]\}$, then $\chi[\mathbf{u}] \subseteq E$.*

Proof. Please see Appendix E. □

The physical interpretation is that $C[\mathbf{u}]$ contains the boundaries of the *cast shadows*. $\chi[\mathbf{u}]$ consists of those edges that cast the shadows. The important (and physically quite intuitive) point here is that every point on the boundary of a cast shadow can be identified with an edge point in $\chi[\mathbf{u}] \subseteq E$. Hence, it is meaningful to talk about the length of the collection of points $\chi[\mathbf{u}]$ that cast shadow edges.

5.2 Perturbation bounds

With all the definitions above, we are ready to show how our bounds are phrased in terms of the extreme values of three physical quantities:

$$\text{Maximum length of shadowing edges:} \quad \chi_\star \doteq \sup_{\mathbf{u} \in \mathbb{S}^2} \text{length}(\chi[\mathbf{u}]), \quad (5.8)$$

$$\text{Minimum visibility:} \quad \nu_\star \doteq \inf_{\mathbf{x} \in \partial \mathcal{O}} \tilde{\nu}(\mathbf{x}) \geq 0, \quad (5.9)$$

$$\text{Maximum albedo:} \quad \rho_\star \doteq \sup_{\mathbf{x} \in \partial \mathcal{O}} \rho(\mathbf{x}) \leq 1. \quad (5.10)$$

For convex objects, $\nu_\star = 1$ and $\chi_\star = 0$. For general objects, $1 - \nu_\star$ and χ_\star can be interpreted as measures of nonconvexity. In terms of these quantities, we obtain perturbation bounds on $\bar{\mathcal{D}}$, \mathcal{T} and \mathcal{P} , which can be combined to bound the error in approximating $\bar{\mathbf{y}}[\mathbf{u}]$:

Theorem 5.3 (Perturbation of direct illumination). *Suppose that \mathcal{O} is a triangulated object, and $\rho(\mathbf{x}) : \partial\mathcal{O} \rightarrow (0, 1]$ is strictly positive. Let $\bar{\mathcal{D}}[\mathbf{u}] \in L^2[\partial\mathcal{O}]$ be as in (4.11). Then for all $\mathbf{u}, \mathbf{u}' \in \mathbb{S}^2$ with $\|\mathbf{u} - \mathbf{u}'\|_2 \leq \sqrt{2}$, we have*

$$\|\bar{\mathcal{D}}[\mathbf{u}] - \bar{\mathcal{D}}[\mathbf{u}']\|_{L^2}^2 \leq 2\rho_\star^2 \text{area}(\partial\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2^2 + 32\sqrt{2}\rho_\star^2 \text{diam}(\mathcal{O}) \chi_\star \|\mathbf{u} - \mathbf{u}'\|_2. \quad (5.11)$$

If \mathcal{O} is convex, we have the tighter bound

$$\|\bar{\mathcal{D}}[\mathbf{u}] - \bar{\mathcal{D}}[\mathbf{u}']\|_{L^2}^2 \leq \rho_\star^2 \text{area}(\partial\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2^2. \quad (5.12)$$

The first term of (5.11) accounts for continuous changes induced by $\langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle_+$. The second term accounts for nonsmooth changes due to cast shadows, which are reflected in the term $\nu(\mathbf{x}, \mathbf{u})$. The proof of Theorem 5.3 is somewhat technical. We delay it to Appendix F.

After direct illumination, the object is also subject to interreflection, \mathcal{T} . This operator is also bounded:

Lemma 5.4. *The operator \mathcal{T} satisfies $\|\mathcal{T}\|_{L^2 \rightarrow L^2} \leq \rho_\star \cdot (1 - \nu_\star)$.*

We prove this bound in Appendix G. In practice, ν_\star is bounded away from zero, and ρ_\star is bounded away from one. This implies that $\|\mathcal{T}\|_{L^2 \rightarrow L^2} < 1$, and

$$\|(\mathcal{I} - \mathcal{T})^{-1}\|_{L^2 \rightarrow L^2} \leq (1 - \rho_\star \cdot (1 - \nu_\star))^{-1}. \quad (5.13)$$

This inequality controls the total effect of interreflection for all bounces. For convex objects, $\nu_\star = 1$, and \mathcal{T} does not participate in the image formation process.

At last, the effect of projection and sampling can be controlled in terms of the properties of the sensor:

Lemma 5.5. *Under our sensor model, let $\ell = \min \{\langle \mathbf{e}_3, \mathbf{x} \rangle \mid \mathbf{x} \in \mathcal{O}\}$ be the depth of the object, and set $\beta_c = \frac{\gamma_c}{4} \frac{d_c^2}{f_c^2}$. The projection and sampling operator \mathcal{P} satisfies*

$$\|\mathcal{P}\|_{L^2 \rightarrow \ell^2} \leq 2^{1/4} \beta_c f_c s_c / \ell. \quad (5.14)$$

Finally, putting these three bounds together, we obtain a perturbation for the images $\bar{\mathbf{y}}[\cdot]$ of \mathcal{O} under point illumination below:

Theorem 5.6. *Under our hypotheses, for any $\mathbf{u}, \mathbf{u}' \in \mathbb{S}^2$ with $\|\mathbf{u} - \mathbf{u}'\|_2 \leq \sqrt{2}$,*

$$\|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2 \leq 2^{1/4} \frac{\beta_c \rho_\star f_c s_c}{\ell(1 - \rho_\star(1 - \nu_\star))} \times \left(2 \text{area}(\partial\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2^2 + 32\sqrt{2} \text{diam}(\mathcal{O}) \chi_\star \|\mathbf{u} - \mathbf{u}'\|_2 \right)^{1/2},$$

Number of Sample Images. To our knowledge, Theorems 5.3-5.6 are new, and could be useful for other problems in vision and graphics. This bound depends only on properties of the object and imaging system that can be known or estimated. In conjunction with Lemma 3.3, it gives a guideline for choosing the sampling density that *guarantees* a representation that works for every illumination $f \in \mathcal{F}_\alpha$.

In particular, as $\|\mathbf{u} - \mathbf{u}'\|_2 \rightarrow 0$, Theorem 5.6 suggests that $\|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2$ is proportional to $\|\mathbf{u} - \mathbf{u}'\|_2^{1/2}$. We can deduce that for guaranteed ε -approximation verification with ambient illumination level α , it would require

$$n(\alpha, \varepsilon) = \frac{\text{const}(\text{sensor}, \text{object})}{(\alpha \varepsilon)^4} \quad (5.15)$$

sample images – polynomial in the approximation error ε , ambient level α and dimension m . This is possible due to the very special structure of the extreme rays $\bar{\mathbf{y}}[\mathbf{u}]$ of C_α : they lie on a submanifold of dimension 2. In contrast, general cone approximation in \mathbb{R}^m requires a number of samples exponential in m [BI76].

6 Cone Preserving Complexity Reduction

Although the sample complexity $n(\alpha, \varepsilon)$ in (5.15) is polynomial in ε^{-1} , it can still be very large. This makes working directly with the dictionary $\bar{\mathbf{A}} \in \mathbb{R}^{m \times n}$ problematic in practice. Hence, we would like to find a surrogate $\hat{\mathbf{A}}$ that is structured in such a way as to enable efficient computation, while still belonging to the set

$$\Omega_0 \doteq \left\{ \hat{\mathbf{A}} \in \mathbb{R}^{m \times n} \mid \delta(\text{cone}(\bar{\mathbf{A}}), \text{cone}(\hat{\mathbf{A}})) \leq \gamma \right\} \quad (6.1)$$

for guaranteed verification. If $\hat{\mathbf{A}}$ can be expressed as $\mathbf{L} + \mathbf{S}$, where \mathbf{L} has rank r and \mathbf{S} has k nonzero entries, product $\hat{\mathbf{A}}\mathbf{x}$ can be computed in time $O((m+n)r+k)$, much smaller than $O(mn)$. Empirical evidence suggests that this model gives a good approximation for images under varying illuminations (Figure 6): the low-rank term captures the smooth variations [BJ03], while cast shadows are often sparse [WYG⁺09]. The effectiveness of such model has been noted, e.g., in [CLMW11], and exploited for robust photometric stereo by [WGS⁺10].

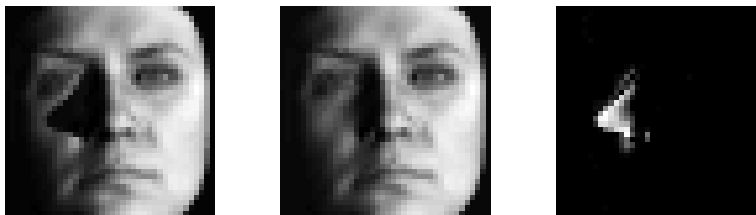


Figure 6: **Low Rank + Sparse Decomposition.** Left: input $\bar{\mathbf{A}}$. Middle: low-rank term \mathbf{L} . Right: sparse term \mathbf{S} .

To build a framework for complexity reduction with guaranteed approximation quality, we start with the following problem, which seeks the *lowest-complexity* pair (\mathbf{L}, \mathbf{S}) that suffice for guaranteed

verification:

$$\begin{aligned} \min_{(\mathbf{L}, \mathbf{S})} \quad & \text{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0 \\ \text{s.t.} \quad & \mathbf{L} + \mathbf{S} = \hat{\mathbf{A}} \in \Omega_0. \end{aligned} \quad (6.2)$$

Note that the constraint illustrates a basic difference between our setting here and all of the aforementioned works on low-rank and sparse recovery. Previous works aimed at statistical estimation of \mathbf{L} and \mathbf{S} , and hence worked with simple constraints of the form $\|\mathbf{L} + \mathbf{S} - \bar{\mathbf{A}}\|_F \leq \varepsilon$. Here, we care about preserving the performance guarantee for detection – in particular, ensuring that $\text{cone}(\mathbf{L} + \mathbf{S})$ and $\text{cone}(\bar{\mathbf{A}})$ are close in Hausdorff sense. This forces us to work with a more complicated set Ω_0 of matrices, giving a very different optimization problem. The following result shows how to convexify Ω_0 , to obtain a tractable convex optimization problem whose solution is guaranteed to well-approximate $\text{cone}(\bar{\mathbf{A}})$, *in the sense required by the application*.

One immediate relaxation is to replace nonconvex objectives rank and ℓ^0 -norm with their convex envelope nuclear norm $\|\cdot\|_*$ (sum of all singular values) and the ℓ^1 -norm (sum of absolute values of all entries) respectively. For the nonconvex domain Ω_0 , we will instead work on one of its convex subsets Ω_1 , which is defined via a bound on a supremum of convex functions of $\hat{\mathbf{A}}$ as follows:

Lemma 6.1. *If $\gamma' \leq \frac{\gamma}{\gamma+1}$, we have $\Omega_1 \subseteq \Omega_0$, where*

$$\Omega_1 \doteq \left\{ \hat{\mathbf{A}} \in \mathbb{R}^{m \times n} \mid \max_{\mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq 1} \|\bar{\mathbf{A}}\mathbf{x} - \hat{\mathbf{A}}\mathbf{x}\|_2 \leq \gamma' \right\}. \quad (6.3)$$

While Ω_1 is convex, it does not have a tractable description, because of the quadratic maximization involved. We use a standard lifting trick, writing $\mathbf{X} = \mathbf{x}\mathbf{x}^T \succeq \mathbf{0}$, to relax this quadratic maximization to a (convex) linear maximization over a semidefinite unknown \mathbf{X} . This gives an upper bound on the maximum in (6.3), giving another convex subset $\Omega_2 \subseteq \Omega_1$, which *does* admit a tractable representation:

Lemma 6.2. *Consider*

$$\Omega_2 \doteq \left\{ \hat{\mathbf{A}} \in \mathbb{R}^{m \times n} \mid \max_{\mathbf{X} \in \mathcal{X}} \langle (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}), \mathbf{X} \rangle \leq (\gamma')^2 \right\},$$

where $\mathcal{X} \doteq \left\{ \mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} \succeq \mathbf{0}, \mathbf{X} \preceq \mathbf{0}, \langle \bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{X} \rangle \leq 1 \right\}$. Then we have $\Omega_2 \subseteq \Omega_1$.

Finally, we reformulate Ω_2 via the dual problem of $\max_{\mathbf{X} \in \mathcal{X}} \langle (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}), \mathbf{X} \rangle$:

Theorem 6.3. *Let $(\mathbf{L}^*, \mathbf{S}^*)$ solve*

$$\begin{aligned} \min_{(\mathbf{L}, \mathbf{S}, \boldsymbol{\mu})} \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{I} & \mathbf{L} + \mathbf{S} - \bar{\mathbf{A}} \\ (\mathbf{L} + \mathbf{S} - \bar{\mathbf{A}})^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0}, \quad \boldsymbol{\mu} \geq \mathbf{0}. \end{aligned} \quad (6.4)$$

with $\bar{\gamma} = (\gamma')^2 \leq (\frac{\gamma}{1+\gamma})^2$, then $\delta(\text{cone}(\bar{\mathbf{A}}), \text{cone}(\mathbf{L}^* + \mathbf{S}^*)) \leq \gamma$.

In contrast to existing matrix decompositions (e.g., [CSPW11, CLMW11]), which aim at statistical estimation, and measure quality of approximation in Frobenius norm, we guarantee approximation in Hausdorff distance $\delta(\cdot, \cdot)$. This is precisely the measure required for worst case verification. We call (6.4) a *cone-preserving low-rank and sparse decomposition*. It can be computed efficiently using the Linearized Alternating Direction Method of Multipliers (L-ADMM) ([ZBBO10, MXZ12]), which converges globally with rate $O(1/k)$ [HY12]. We describe the numerical implementation and convergence theory associated with this approach in more detail in Appendix J.

7 Numerical Experiment

We render images from 3D triangulated object models following a simplified imaging process $\mathbf{y}[f] = \mathcal{PD}[f]$. Thus, our simulations include cast shadows, but not interreflection.⁹ Camera parameters $\gamma_c = f_c = d_c = 1$ and $s_c = 0.003$ are fixed throughout our experiments.

Verifying the Perturbation Bound. We compare the bound in Theorem 5.6, denoted as $\text{PerturbationBound}(\mathbf{u}, \mathbf{u}')$, to the actual difference $\|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2$ for three different object shapes. The maximum ratio between those two quantities can be expressed as

$$r \doteq \max_{\mathbf{u}, \mathbf{u}' \text{ adjacent}} \left\{ \frac{\|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2}{\text{PerturbationBound}(\mathbf{u}, \mathbf{u}')} \right\} \quad (7.1)$$

In our experiment, the set of point illuminations is generated using a uniform grid, $\theta = \pi/360, 2\pi/360, \dots$ and $\phi = \pi/360, 2\pi/360, \dots$ in spherical coordinates. Results are listed in Table 1: the ratio is always bounded by 1, corroborating our theoretical results.

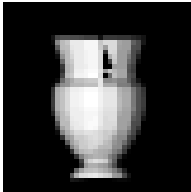

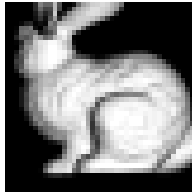
Object	Vase	Face	Bunny
			
r	0.1809	0.0302	0.0290

Table 1: **Tightness of the bounds.** Largest ratio r between experimental observation and theoretical upper bound for three different objects. The bound holds in all cases, and is tightest for the vase.

Order of Perturbation Bound. We next consider the behavior of our bounds when $\|\mathbf{u} - \mathbf{u}'\|_2 \rightarrow 0$. Our bounds predict that in the worst case, the change in irradiance $\bar{\mathcal{D}}[\mathbf{u}]$ should be proportional

⁹This approximation neglects the nontrivial interreflection terms, $\mathcal{T} + \mathcal{T}^2 + \dots$. These terms are at most on the order of ρ_* , the maximum albedo. The approximation $\mathbf{y}[f] = \mathcal{PD}[f]$ can be (loosely) considered to be the limiting case as ρ_* becomes small. Here, we make this approximation to make it easier to efficiently simulate the imaging process.

to $\|\mathbf{u} - \mathbf{u}'\|_2^{1/2}$. We investigate this using a toy object composed of two perpendicular surfaces S_1 and S_2 shown in Figure 7 with \mathbf{u} fixed, and \mathbf{u}' changing slowly. Figure 7 (right) shows how $\|\bar{\mathcal{D}}[\mathbf{u}'] - \bar{\mathcal{D}}[\mathbf{u}]\|$ depends on $\|\mathbf{u} - \mathbf{u}'\|_2$. Both the theoretical prediction and the computed value appear to be proportional to $\|\mathbf{u} - \mathbf{u}'\|_2^{1/2}$.¹⁰ This suggests that in the worst case, our theory may be tight up to constant factors. The theoretical prediction curve neglects constants, and simply draws $\text{diam}(\mathcal{O}) \times \chi_\star \times \sqrt{\|\mathbf{u} - \mathbf{u}'\|_2}$.

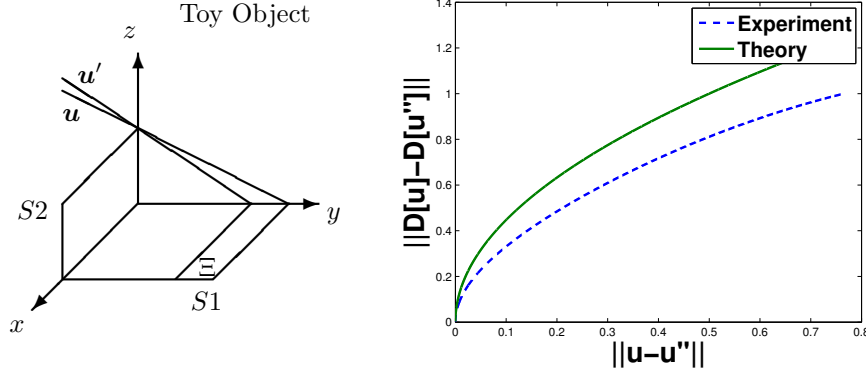


Figure 7: **Order of Perturbation Bound.** Here, in both theory and simulation the change in $\bar{\mathcal{D}}[\mathbf{u}]$ is proportional to $\|\mathbf{u} - \mathbf{u}'\|^{1/2}$.

Cone Preserving Complexity Reduction. We demonstrate the ability of our solution to (6.4) to reduce the complexity of the representation, while preserving the conic hull. We start with $n = 648$ images of a face under point illuminations, with resolution 40×40 . We solve the low-rank and sparse cone approximation problem in Theorem 6.3 for varying cone distances γ with λ simply chosen as $\sqrt{\max(m, n)}$. Figure 8 plots the ratio complexity of $\hat{\mathbf{A}}$ and \mathbf{A} , or $\frac{(m+n)r+s}{mn}$, where r is the rank of the recovered low-rank term and s is the number of nonzero entries in the recovered sparse term. The decomposition reduces the complexity in all cases; the reduction becomes more pronounced as α increases. This is expected, since the cone C_α becomes smaller as α increases. This reduction in complexity suggests that although the number of extreme rays is large, there may be quite a bit of additional structure *across* the extreme rays, that could be exploited by more sophisticated cone constructions.

Application Sketch. To conceptually justify the advantage of our cone approximation methodology in verification under poor illumination conditions, we compare the receiver operating characteristic (ROC) curves for 5 verification dictionaries obtained from same 3D face model under ambient level $\alpha = 0.1$: convex cone C_1 composed of 2592 images, corresponds to the ε -approximation of the original illumination cone; C_2 is the γ -approximation of C_1 with $L + S$ structure ($\gamma = 0.11$); C_3 is rendered under 19 illumination directions corresponding to subsets 1 and 2 of Yale B [GBK01] (roughly, the setting of [WYG⁺09]); C_4 is rendered under all 64 illumination directions considered

¹⁰In Figure 7, we rescale the theoretical prediction for clearer comparison – our goal is only to show that the exponent is $1/2$.

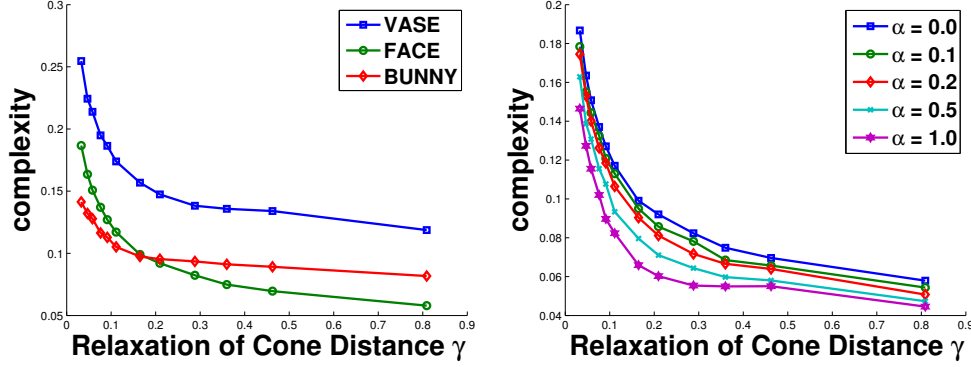


Figure 8: **Cone Complexity Reduction** for different nonconvex objects under zero ambient level ($\alpha = 0$) (left) and for *face* under different ambient illumination levels (right).

in [GBK01]. Finally, motivated by [BJ03], we also consider the subspace S spanned by 9 principal components of C_1 .

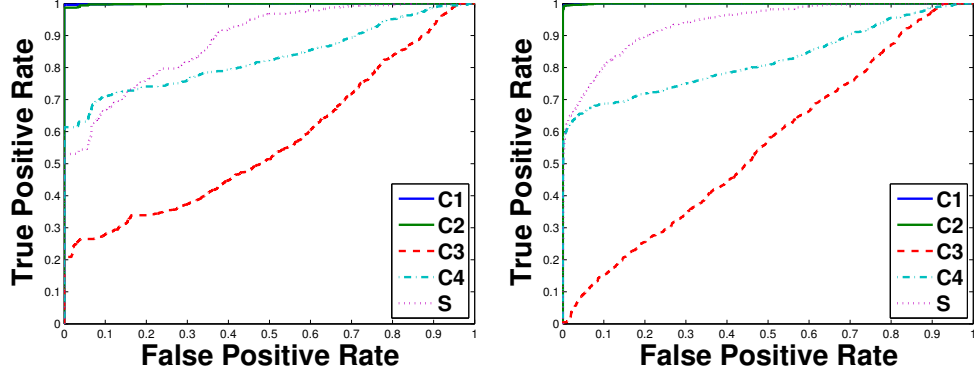


Figure 9: **ROC Curves** for different dictionaries, with test images under uniform random illumination (left) and extreme illumination (right). The dictionaries are C_1 : ε approximation, C_2 : low-rank and sparse approximation, C_3 - C_4 : point illuminations distributed similar to [GBK01], S : nine-dimensional linear subspace.

Our test data consist of 1,000 positive images under 1,000 illumination directions and 3,000 negative images of 3 other subjects. We consider two distributions for the illumination directions – uniform on the sphere (roughly corresponding to the “average case”), and uniform on the set of $\mathbf{u} \in \mathbb{S}^2$ for which $-0.1 \leq u_3 \leq 0.4$. Here, the u_3 axis is the camera axis. Arguably, the second set is more challenging. Figure 9 shows the ROC for a simple verification test based on the distance to the models. As suggested by our theoretical results, both C_1 and C_2 perform almost perfectly. The simpler models C_3 , C_4 , S perform better than chance, but still break down frequently. We view this result as illustrating a tradeoff in illumination representation: uniformly good performance is possible, *if* we can afford a more complex representation. The cone preserving low-rank and sparse

decomposition gives a way to control the complexity of the representation, while still maintaining this good performance.

8 Discussion

There are several directions for future work. Although our cone construction scheme guarantees worst case verification, the number of samples required is likely to be very large, in particular for small ε : when $\varepsilon = 0.01$, our theorem requires about 10^{25} images under ambient level $\alpha = 1$. Our experiments on complexity reduction suggest that there should exist a simpler representation, if we can take advantage of the structure of shadows.

To use the results in a practical recognition system, we need to account for variations in object pose as well. This can be done using local optimization heuristics, or simply building models at a set of reference poses [GBK01]. It will be important to have very concise models for each pose; the complexity reduction by convex programming is one means of achieving this.

Here, we have considered object instance *verification*, rather than object instance *recognition*. The “yes/no” question in verification forces us to confront basic questions about the set of images of the object. Nevertheless, we believe our methodology will be useful for recognition as well. For example, one could build models \hat{C} for each class and assigning the test sample to the closest model in angle. For recognition problems, the formulations and goals for sampling and complexity reduction may also change.

We anticipate three classes of practical application of our results. The first is in instance detection/recognition using 3D models and 2D test images. The second is in instance detection/recognition with active acquisition of training data, e.g., in face recognition for access control [WWG⁺12]. The final, more speculative application is in instance detection/recognition with large families of objects with similar gross shape and appearance. In face recognition, learned models for physical variabilities (albedo and illumination) are often used in conjunction with deformable models [CET01]. In many practical settings, this approach mitigates the difficulties associated with small training datasets – they can work with as few as one image [WZL⁺, ZYZ⁺13]. Our results could give a way of learning a set of canonical illumination models, which capture effects such as cast shadows, and which could be adapted to each new input subject.

Acknowledgment

It is a great pleasure to acknowledge conversations with Donald Goldfarb (Columbia), Yi Ma (MSRA), Stefano Soatto (UCLA), Andrew Wagner (K.U. Leuven), Bin Yu (Berkeley), Zhengdong Zhang (MIT), Zihan Zhou (Penn. State). JW also gratefully acknowledges support from Columbia University and the Office of Naval Research.

References

- [BI76] E. M. Bronshteyn and L. D. Ivanov. The approximation of convex sets by polyhedra. *Siberian Mathematical Journal*, 16(5):852–853, 1976.

- [BJ03] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 25(2):218–233, 2003.
- [BK98] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision (IJCV)*, 28(3):245–260, 1998.
- [CET01] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):681–685, 2001.
- [CLMW11] E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the Association for Computing Machinery*, 58(3), 2011.
- [Con90] J. B. Conway. *A Course in Functional Analysis*. Springer, 1990.
- [Cos00] M. Coste. *An Introduction to Semialgebraic Geometry*. Istituti Editoriali e Poligrafici Internazionali, 2000.
- [CSPW11] V. Chandrasekaran, S. Sanghavi, P. Parillo, and A. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.
- [CYZ⁺05] T. Chen, W. Yin, X. S. Zhou, D. Domaniciu, , and T. S. Huang. Illumination normalization for face recognition and uneven background correction using total variation based image models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [EHY95] R. Epstein, P.W. Hallinan, and A.L. Yuille. 5 ± 2 eigenimages suffice: an empirical investigation of low-dimensional lighting models. In *Physics-Based Modeling in Computer Vision*, 1995.
- [FSB04] D. Frolova, D. Simakov, and R. Basri. Accuracy of spherical harmonic approximations for images of lambertian objects under far and near lighting. In *European Conference on Computer Vision (ECCV)*, 2004.
- [GBK01] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 23(6):643–660, 2001.
- [Hor86] B. K. Horn. *Robot Vision*. McGraw-Hill Higher Education, 1986.
- [HY12] B. He and X. Yuan. On the $o(1/n)$ convergence rate of the douglas-rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.
- [LHK05] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 27(5):684–698, 2005.
- [Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

- [MLJ09] X. Mei, H. Ling, and D. W. Jacobs. Sparse representation of cast shadows via l1-regularized least squares. In *IEEE International Conference on Computer Vision (ICCV)*, 2009.
- [MXZ12] S. Ma, L. Xue, and H. Zou. Alternating direction methods for latent variable gaussian graphical model selection. *arXiv preprint arXiv:1206.1275*, 2012.
- [Nes07] Y. Nesterov. Gradient methods for minimizing composite objective function. *Center for Operations Research and Econometrics*, 2007.
- [Ram02] R. Ramamoorthi. Analytic pca construction for theoretical analysis of lighting variability in images of a lambertian object. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 24(10):1322–1333, 2002.
- [RKB] R. Ramamoorthi, M. Koudelka, and P. Belhumeur. A fourier theory for cast shadows. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 27(2):288–295.
- [SRR01] A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 23(2):129–139, 2001.
- [SSTB12] A. Savran, B. Sankur, and M. Taha Bilge. Comparative evaluation of 3d vs. 2d modality for automatic detection of facial action units. *Pattern Recognition*, 45(2):767–782, 2012.
- [WGS⁺10] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision (ACCV)*, 2010.
- [WWG⁺12] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 34(2):372–386, 2012.
- [WYG⁺09] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 31(2):210–227, 2009.
- [WZL⁺] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI)*, 31(11):1968–1984.
- [ZBBO10] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.
- [ZBO11] X. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011.
- [ZYZ⁺13] L. Zhuang, A. Yang, Z. Zhou, S. Sastry, and Y. Ma. Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. In *Computer Vision and Pattern Recognition (CVPR)*, 2013.

Appendix

A Proof of Lemma 2.3

Proof. For all \mathbf{y} of norm one such that $\angle(\mathbf{y}, C) \leq \tau$, $\exists \bar{\mathbf{y}} \in \text{cl}(C)$ with $\|\mathbf{y} - \bar{\mathbf{y}}\| \leq \sin \tau$. Moreover, $\|\bar{\mathbf{y}}\|_2 \leq 1$. By (2.6), we know that $\exists \hat{\mathbf{y}} \in \text{cl}(\hat{C})$ s.t. $\|\bar{\mathbf{y}} - \hat{\mathbf{y}}\| \leq \delta(C, \hat{C})$. By triangle inequality, $\|\mathbf{y} - \hat{\mathbf{y}}\| \leq \sin \tau + \delta(C, \hat{C})$, which implies that $\angle(\mathbf{y}, \hat{C}) \leq \text{asin}(\sin \tau + \delta(C, \hat{C})) \leq \xi$. Therefore $\mathfrak{D}_{\xi}^{\hat{C}}(\mathbf{y}) = \text{ACCEPT}$.

Conversely, if $\mathfrak{D}_{\xi}^{\hat{C}}(\mathbf{y}) = \text{ACCEPT}$, then $\angle(\mathbf{y}, \hat{C}) \leq \xi$. If $\|\mathbf{y}\|_2 \leq 1$, this implies that there exists $\hat{\mathbf{y}} \in \text{cl}(\hat{C})$ of norm at most one such that $\|\hat{\mathbf{y}} - \mathbf{y}\|_2 \leq \sin(\xi)$. Moreover, from the definition of δ , there exists $\bar{\mathbf{y}} \in C \cap \mathfrak{B}(0, 1)$ such that $\|\bar{\mathbf{y}} - \hat{\mathbf{y}}\|_2 \leq \delta(C, \hat{C})$. By the triangle inequality,

$$\|\mathbf{y} - \bar{\mathbf{y}}\| \leq \sin(\xi) + \delta(C, \hat{C}).$$

Moreover, if

$$\text{asin}(\sin(\xi) + \delta(C, \hat{C})) \leq (1 + \eta)\tau, \quad (\text{A.1})$$

we have $\angle(\mathbf{y}, C) \leq (1 + \eta)\tau$. Hence, whenever (A.1) holds, for every \mathbf{y} such that $\mathfrak{D}_{\xi}^{\hat{C}}(\mathbf{y}) = \text{ACCEPT}$, we have $\angle(\mathbf{y}, C) \leq (1 + \eta)\tau$, and hence $\mathfrak{D}_{\xi}^{\hat{C}}(\mathbf{y}) = \text{REJECT}$ for all \mathbf{y} with $\angle(\mathbf{y}, C) > (1 + \eta)\tau$. This condition holds whenever

$$\xi \leq \text{asin}(\sin(\tau + \eta\tau) - \delta(C, \hat{C})). \quad (\text{A.2})$$

Take together with the first paragraph, this condition and $\xi \geq \text{asin}(\sin \tau + \delta(C, \hat{C}))$ imply that $\mathfrak{D}_{\xi}^{\hat{C}} \in \hat{\mathbb{D}}_{\tau, \eta}^C$, which establishes our claim. \square

B Proofs from Section 3

The definition of the Riemann integral on \mathbb{R}^2 gives the following:

Lemma B.1. *Let $\mathbf{h} : \mathbb{S}^2 \rightarrow \mathbb{R}^m$ be a vector-valued function which is nonnegative and Riemann integrable. Then for every $\varepsilon > 0$, there exists $N \in \mathbb{Z}_+$, $\mathbf{u}_1 \dots \mathbf{u}_N \in \mathbb{S}^2$, and $\lambda_1 \dots \lambda_N \geq 0$ such that*

$$\left\| \int_{\mathbf{u} \in \mathbb{S}^2} \mathbf{h}[\mathbf{u}] d\mathbf{u} - \sum_{i=1}^N \lambda_i \mathbf{h}[\mathbf{u}_i] \right\|_2 \leq \varepsilon. \quad (\text{B.1})$$

Proof. Again letting $W = [0, 2\pi] \times [0, \pi]$, and let $\eta(\theta, \phi) = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi)$ denote the spherical coordinate map. Consider a single coordinate j . From the definition of the Riemann integral, we have

$$\int_{\mathbf{u} \in \mathbb{S}^2} h_j[\mathbf{u}] d\mathbf{u} = \int_W h_j[\eta(\theta, \phi)] \sin \phi d(\theta, \phi), \quad (\text{B.2})$$

where the right hand side is a Riemann integral on \mathbb{R}^2 . For every $\eta > 0$, there exists a partition Π_j of $[0, 2\pi] \times [0, \pi]$ such that

$$\sum_{R \in \Pi_j} \left[\sup_{(\theta, \phi) \in R} \tilde{\mathbf{h}}_j[\eta(\theta, \phi)] \sin \phi \right] \text{vol}(R) - \eta \leq \int_{\mathbf{u}} \mathbf{h}_j[\mathbf{u}] d\mathbf{u} \leq \sum_{R \in \Pi_j} \left[\inf_{(\theta, \phi) \in R} \tilde{\mathbf{h}}_j[\eta(\theta, \phi)] \sin \phi \right] \text{vol}(R) + \eta.$$

Choose such a partition Π_j for each j , and let $\Pi = \{R_1, \dots, R_L\}$ be the common refinement. Then for any choice of $(\theta_i, \phi_i) \in R_i$, we have

$$\left\| \sum_{i=1}^L \mathbf{h}_j[\eta(\theta_i, \phi_i)] \sin \phi_i \times \text{vol}(R_i) - \int_{(\theta, \phi)} \mathbf{h}[\eta(\theta, \phi)] \sin \phi d(\theta, \phi) \right\|_{\infty} \leq \eta. \quad (\text{B.3})$$

For all $i \in I$, set $\mathbf{u}_i = \eta(\theta_i, \phi_i)$, and $\lambda_i = \text{vol}(R_i) \sin \phi_i$. Then

$$\left\| \sum_{i=1}^L \lambda_i \mathbf{h}[\mathbf{u}_i] - \int_{\mathbf{u}} \mathbf{h}[\mathbf{u}] d\mathbf{u} \right\|_{\infty} \leq \eta. \quad (\text{B.4})$$

Setting $\eta = \varepsilon/\sqrt{m}$, we obtain the result. \square

Proof of Lemma 3.1. Below, we prove Lemma 3.1.

Proof. For notational convenience, let $\Psi = \text{cone}(\{\bar{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\})$. Consider $\mathbf{y}_0 \in C_0$. Then $\mathbf{y}_0 = \mathbf{y}[f] = \int_{\mathbf{u}} f(\mathbf{u}) \bar{\mathbf{y}}[\mathbf{u}] d\mathbf{u}$, for some nonnegative, Riemann integrable f . The vector valued function $\bar{\mathbf{y}}[\mathbf{u}]f(\mathbf{u})$ is Riemann integrable. By Lemma B.1, for every $\varepsilon > 0$, there exists $\hat{\mathbf{y}}_{\varepsilon} \in \Psi$ with $\|\mathbf{y}_0 - \hat{\mathbf{y}}_{\varepsilon}\|_2 \leq \varepsilon$. Hence,

$$\mathbf{y}_0 \in \text{cl}(\Psi), \quad (\text{B.5})$$

and $C_0 \subseteq \text{cl}(\Psi)$.

We complete the proof by showing that $\Psi \subseteq \text{cl}(C_0)$. By continuity of $\bar{\mathbf{y}}[\cdot]$, for any $\mathbf{u}_0 \in \mathbb{S}^2$, and any $\varepsilon > 0$, there exists $\eta > 0$ such that

$$\|\bar{\mathbf{y}}[\mathbf{u}_0] - \bar{\mathbf{y}}[\mathbf{v}]\|_2 \leq \varepsilon \quad \forall \mathbf{v} \in \mathfrak{B}(\mathbf{u}_0, \eta) \cap \mathbb{S}^2, \quad (\text{B.6})$$

where $\mathfrak{B}(\mathbf{u}, r)$ is the ℓ^2 ball of radius r around \mathbf{u} . Let $f_{\varepsilon} : \mathbb{S}^2 \rightarrow \mathbb{R}$ via

$$f_{\varepsilon}(\mathbf{v}) = \frac{1}{\text{area}(\mathfrak{B}(\mathbf{u}_0, \eta) \cap \mathbb{S}^2)} \mathbb{1}_{\|\mathbf{v} - \mathbf{u}_0\|_2 \leq \eta}. \quad (\text{B.7})$$

Then f_{ε} is Riemann integrable, and

$$\begin{aligned} \|\mathbf{y}[f_{\varepsilon}] - \bar{\mathbf{y}}[\mathbf{u}_0]\|_2 &= \left\| \frac{1}{\text{area}(\mathfrak{B}(\mathbf{u}_0, \eta) \cap \mathbb{S}^2)} \int_{\|\mathbf{u} - \mathbf{u}_0\| \leq \eta} \bar{\mathbf{y}}[\mathbf{u}] d\mathbf{u} - \bar{\mathbf{y}}[\mathbf{u}_0] \right\|_2 \\ &\leq \frac{1}{\text{area}(\mathfrak{B}(\mathbf{u}_0, \eta) \cap \mathbb{S}^2)} \int_{\|\mathbf{u} - \mathbf{u}_0\| \leq \eta} \|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}_0]\|_2 d\mathbf{u} \\ &\leq \frac{1}{\text{area}(\mathfrak{B}(\mathbf{u}_0, \eta) \cap \mathbb{S}^2)} \int_{\|\mathbf{u} - \mathbf{u}_0\| \leq \eta} \varepsilon d\mathbf{u} \\ &\leq \varepsilon. \end{aligned} \quad (\text{B.8})$$

Since this is true for every $\varepsilon > 0$, $\bar{\mathbf{y}}[\mathbf{u}] \in \text{cl}(C_0)$, and so $\Psi \subseteq \text{cl}(C_0)$, completing the proof. \square

Proof of Lemma 3.2. Below, we prove Lemma 3.2.

Proof. For $f = \alpha\omega + f_d \in \mathcal{F}_a$, write

$$\mathbf{y}[f] = \int_{\mathbf{u}} \bar{\mathbf{y}}[\mathbf{u}] f(\mathbf{u}) d\mathbf{u} = \int_{\mathbf{u}} \bar{\mathbf{y}}[\mathbf{u}] (\alpha\omega(\mathbf{u}) + f_d(\mathbf{u})) d\mathbf{u} \quad (\text{B.9})$$

$$= \alpha \mathbf{y}_a + \int_{\mathbf{u}} \bar{\mathbf{y}}[\mathbf{u}] f_d(\mathbf{u}) d\mathbf{u}, \quad (\text{B.10})$$

$$= \alpha (1 - \|f_d\|_{L^1}) \mathbf{y}_a + \int_{\mathbf{u}} \check{\mathbf{y}}[\mathbf{u}] f_d(\mathbf{u}) d\mathbf{u}. \quad (\text{B.11})$$

Repeating arguments in the proof of Lemma 3.1, and using that $\check{\mathbf{y}}[\mathbf{u}]$ is continuous in \mathbf{u} , we have

$$\int_{\mathbf{u}} \check{\mathbf{y}}[\mathbf{u}] f_d(\mathbf{u}) d\mathbf{u} \in \text{cl}(\text{cone}(\{\check{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\})). \quad (\text{B.12})$$

Hence, from (B.11), $\|f_d\|_{L^1} \leq 1$, and the fact that

$$\mathbf{y}_a \in \text{cl}(\text{cone}(\{\check{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\})) \quad (\text{B.13})$$

we have $\mathbf{y}[f] \in \text{cl}(\text{cone}(\{\check{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\}))$, and so $C_\alpha \subseteq \text{cl}(\check{C})$.

Conversely, repeating arguments of Lemma 3.1, we can show that the generators $\{\check{\mathbf{y}}[\mathbf{u}] \mid \mathbf{u} \in \mathbb{S}^2\}$ and \mathbf{y}_a are all elements of $\text{cl}(C_\alpha)$, and hence $\check{C} \subseteq \text{cl}(C_\alpha)$, completing the proof. \square

Proof of Lemma 3.3.

Proof. Set

$$\varepsilon = \sup_{\mathbf{u} \in \mathbb{S}^2} \min_i \left\| \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}_i]}{\|\check{\mathbf{y}}[\mathbf{u}_i]\|_2} \right\|_2, \quad (\text{B.14})$$

and let \mathbf{w}_\star realize the supremum in the definition of

$$\eta_\star = \sup_{\|\mathbf{w}\|_2 \leq 1} \inf_i \left\langle \mathbf{w}, \frac{\check{\mathbf{y}}[\mathbf{u}_i]}{\|\check{\mathbf{y}}[\mathbf{u}_i]\|_2} \right\rangle. \quad (\text{B.15})$$

Since $\bar{C} \subseteq \check{C}$, we have

$$\delta(\check{C}, \bar{C}) = \sup_{\mathbf{y} \in \check{C} \setminus \{0\}} \frac{d(\mathbf{y}, \bar{C})}{\|\mathbf{y}\|_2}. \quad (\text{B.16})$$

By Caratheodory's theorem, for every $\mathbf{y} \in \check{C}$, there exist $\mathbf{v}_1 \dots \mathbf{v}_m \in \mathbb{S}^2$ and scalars $\lambda_1 \dots \lambda_m \geq 0$, $\zeta \geq 0$ such that

$$\mathbf{y} = \sum_{j=1}^m \lambda_j \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} + \zeta \mathbf{y}_a. \quad (\text{B.17})$$

For each j , choose \mathbf{u}_{i_j} such that $\left\| \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}_{i_j}]}{\|\check{\mathbf{y}}[\mathbf{u}_{i_j}]\|_2} \right\|_2$ is minimal. Then we have

$$\begin{aligned}
d(\mathbf{y}, \bar{C}) &\leq \left\| \sum_j \lambda_j \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} + \zeta \mathbf{y}_a - \sum_j \lambda_j \frac{\check{\mathbf{y}}[\mathbf{u}_{i_j}]}{\|\check{\mathbf{y}}[\mathbf{u}_{i_j}]\|_2} - \zeta \mathbf{y}_a \right\|_2 \\
&\leq \sum_j \lambda_j \left\| \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}_{i_j}]}{\|\check{\mathbf{y}}[\mathbf{u}_{i_j}]\|_2} \right\|_2 \\
&\leq \sum_j \lambda_j \varepsilon \\
&\leq \frac{\varepsilon}{\eta_\star} \sum_{j=1}^m \lambda_j \left\langle \mathbf{w}_\star, \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} \right\rangle \\
&= \frac{\varepsilon}{\eta_\star} \left\langle \mathbf{w}_\star, \sum_j \lambda_j \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} \right\rangle \\
&\leq \frac{\varepsilon}{\eta_\star} \left\| \sum_j \lambda_j \frac{\check{\mathbf{y}}[\mathbf{v}_j]}{\|\check{\mathbf{y}}[\mathbf{v}_j]\|_2} \right\|_2 \\
&\leq \frac{\varepsilon}{\eta_\star} \|\mathbf{y}\|_2
\end{aligned} \tag{B.18}$$

Hence, $d(\mathbf{y}, \bar{C})/\|\mathbf{y}\|_2 \leq \varepsilon/\eta_\star$. We finish the proof by noting that

$$\begin{aligned}
\left\| \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}']}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} \right\|_2 &\leq \left\| \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} + \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}']}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} \right\|_2 \\
&\leq \left\| \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} \right\|_2 + \left\| \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} - \frac{\check{\mathbf{y}}[\mathbf{u}']}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} \right\|_2 \\
&\leq \|\check{\mathbf{y}}[\mathbf{u}]\|_2 \left| \frac{\|\check{\mathbf{y}}[\mathbf{u}']\|_2 - \|\check{\mathbf{y}}[\mathbf{u}]\|_2}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2 \|\check{\mathbf{y}}[\mathbf{u}]\|_2} \right| + \frac{\|\check{\mathbf{y}}[\mathbf{u}] - \check{\mathbf{y}}[\mathbf{u}']\|_2}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} \\
&\leq 2 \frac{\|\check{\mathbf{y}}[\mathbf{u}] - \check{\mathbf{y}}[\mathbf{u}']\|_2}{\|\check{\mathbf{y}}[\mathbf{u}']\|_2} \\
&\leq \frac{2 \|\check{\mathbf{y}}[\mathbf{u}] - \check{\mathbf{y}}[\mathbf{u}']\|_2}{\alpha \|\mathbf{y}_a\|_2} \\
&= \frac{2 \|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}']\|_2}{\alpha \|\mathbf{y}_a\|_2},
\end{aligned} \tag{B.19}$$

and hence

$$\varepsilon \leq \sup_{\mathbf{u}} \min_i \frac{2 \|\bar{\mathbf{y}}[\mathbf{u}] - \bar{\mathbf{y}}[\mathbf{u}_i]\|_2}{\alpha \|\mathbf{y}_a\|_2}, \tag{B.20}$$

completing the proof. For the bound $\eta_\star \geq 1/\sqrt{m}$, note that if we choose $\mathbf{w} = m^{-1/2} \mathbf{1}$ in the right hand side of (B.15), then since the $\check{\mathbf{y}}[\mathbf{u}]$ are elementwise nonnegative, we for each $\mathbf{u} \in \mathbb{S}^2$

$$\left\langle \mathbf{w}, \frac{\check{\mathbf{y}}[\mathbf{u}]}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} \right\rangle = \frac{1}{\sqrt{m}} \frac{\|\check{\mathbf{y}}[\mathbf{u}]\|_1}{\|\check{\mathbf{y}}[\mathbf{u}]\|_2} \geq \frac{1}{\sqrt{m}}. \tag{B.21}$$

This completes the proof. \square

C Integrating on $\partial\mathcal{O}$

For each triangle

$$\Delta_i = \text{conv} \{z_i, z'_i, z''_i\}, \quad (\text{C.1})$$

we can find an open triangle $U_i \subset \mathbb{R}^2$, and an isometry $\varphi_i : U_i \rightarrow \text{relint}(\Delta_i)$. To make this more concrete, we can let $\mathbf{B}_i \in \mathbb{R}^{3 \times 2}$ be a matrix whose columns are an orthonormal basis for the span $\{z'_i - z_i, z''_i - z_i\}$, write

$$\varphi_i(\mathbf{w}) = z_i + \mathbf{B}_i \mathbf{w}, \quad (\text{C.2})$$

and

$$U_i = \varphi_i^{-1}[\text{relint}(\Delta_i)]. \quad (\text{C.3})$$

We construct an integral $\partial\mathcal{O}$ as follows. For each i , define a σ -algebra Σ_i consisting of all sets of the form $\varphi_i[S]$, where $S \subseteq U_i$ is Lebesgue measurable. Let $\Sigma_{\partial\mathcal{O}}$ be the smallest σ -algebra containing each of the Σ_i . Define a measure $\mu_{\partial\mathcal{O}} : \Sigma_{\partial\mathcal{O}} \rightarrow \mathbb{R}_+$ via

$$\mu_{\partial\mathcal{O}}(S) = \sum_i \mu(\varphi_i^{-1}[S \cap \text{relint}(\Delta_i)]), \quad (\text{C.4})$$

where μ is the Lebesgue measure on \mathbb{R}^2 . It is easy to verify that $\mu_{\partial\mathcal{O}}$ is measure, making $(\Phi, \Sigma_{\partial\mathcal{O}}, \mu_{\partial\mathcal{O}})$ a measure space, with Lebesgue integral

$$\int g(\mathbf{x}) d\mu_{\partial\mathcal{O}}(\mathbf{x}) = \sum_i \int_{U_i} g \circ \varphi_i d\mu. \quad (\text{C.5})$$

This gives an integral over Φ . It extends to an integral over $\partial\mathcal{O}$ as a whole: for $g : \partial\mathcal{O} \rightarrow \mathbb{R}$, we define its integral to be the integral of its restriction to Φ .

D Proof of Lemma 4.2

We prove Lemma 4.2, which writes the imaging map $\mathbf{y}[f] = \mathcal{P}(\mathcal{I} - \mathcal{T})^{-1}\mathcal{D}[f]$ as an integral of the form $\int f(\mathbf{u}) \bar{\mathbf{y}}[\mathbf{u}] d\sigma(\mathbf{u})$, where $\bar{\mathbf{y}}[\mathbf{u}] = \mathcal{P}(\mathcal{I} - \mathcal{T})^{-1}\bar{\mathcal{D}}[\mathbf{u}]$:

Proof. We will show that for any Lebesgue integrable f ,

$$\mathbf{y}[f] = \int f(\mathbf{u}) \bar{\mathbf{y}}[\mathbf{u}] d\sigma(\mathbf{u}). \quad (\text{D.1})$$

By Theorem 5.6, $\bar{\mathbf{y}}[\mathbf{u}]$ is continuous in \mathbf{u} . Since the product of a nonnegative Riemann integrable f and $\bar{\mathbf{y}}[\mathbf{u}]$ is Riemann integrable, this expression is equal to the Riemann integral

$$\mathbf{y}[f] = \int f(\mathbf{u}) \bar{\mathbf{y}}[\mathbf{u}] d\mathbf{u}, \quad (\text{D.2})$$

as desired. To show (D.1), we use Tonelli's theorem and monotone convergence. It is not difficult to show that the Riemann-integrable function f is $\Sigma_{\mathbb{S}^2}$ -measurable, $\bar{\mathcal{D}}[\mathbf{u}](\mathbf{x})$ is $\Sigma_{\mathbb{S}^2} \times \Sigma_{\partial\mathcal{O}}$ measurable,

and κ is $\Sigma_{\partial\mathcal{O}} \times \Sigma_{\partial\mathcal{O}}$ -measurable. By repeated application of Tonelli's theorem, and using that the integrands are nonnegative, we obtain

$$\begin{aligned}\mathcal{T}^i\mathcal{D}[f] &= \int \kappa(\mathbf{x}, \mathbf{x}_i) \dots \left(\int \kappa(\mathbf{x}_2, \mathbf{x}_1) \left(\int \bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}_1) f(\mathbf{u}) d\sigma(\mathbf{u}) \right) d\mu_{\partial\mathcal{O}}(\mathbf{x}_1) \right) \dots d\mu_{\partial\mathcal{O}}(\mathbf{x}_i) \\ &= \int f(\mathbf{u}) \left(\int \kappa(\mathbf{x}, \mathbf{x}_i) \dots \left(\int \kappa(\mathbf{x}_2, \mathbf{x}_1) \bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}_1) d\mu_{\partial\mathcal{O}}(\mathbf{x}_1) \right) \dots d\mu_{\partial\mathcal{O}}(\mathbf{x}_i) \right) d\sigma(\mathbf{u}) \\ &= \int f(\mathbf{u}) (\mathcal{T}^i \bar{\mathcal{D}}[\mathbf{u}])(\mathbf{x}) d\sigma(\mathbf{u}).\end{aligned}\tag{D.3}$$

By monotone convergence,

$$\sum_{i=0}^{\infty} \mathcal{T}^i \mathcal{D}[f](\mathbf{x}) = \int f(\mathbf{u}) \left(\sum_{i=0}^{\infty} \mathcal{T}^i \bar{\mathcal{D}}[\mathbf{u}] \right) (\mathbf{x}) d\sigma(\mathbf{u}).\tag{D.4}$$

One more application of Tonelli's theorem gives (D.1), completing the proof. \square

E Proofs from Section 5.1

Proof of Lemma 5.1.

Proof. Let $\mathbf{x} \in S[\mathbf{u}]^c$. By definition, $\bar{\mathcal{D}}(\mathbf{x}) = 0$ for all $\mathbf{x} \in E$, and so we may assume that $\mathbf{x} \in \Phi = \partial\mathcal{O} \setminus E$. Hence, $\mathbf{x} \in \text{relint}(\Delta)$ for some face Δ . Moreover, the definition of a triangulated object implies that there exists $\tau > 0$ such that $\mathfrak{B}(\mathbf{x}, \tau) \cap \partial\mathcal{O} \subseteq \Delta$.

Suppose, for purposes of contradiction, that there does not exist $r_0 > 0$ such that $\mathfrak{B}(\mathbf{x}, r_0) \cap \partial\mathcal{O} \subseteq S[\mathbf{u}]^c$. Then there exists a sequence $(\mathbf{x}_i)_{i=1}^{\infty} \subset S[\mathbf{u}]$, with $\mathbf{x}_i \rightarrow \mathbf{x}$. By dropping finitely many terms, we may assume $\mathbf{x}_i \in \text{relint}(\Delta)$ for all i . Since $\mathbf{x} \in S[\mathbf{u}]^c$, $\langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle > 0$, and so $\langle \mathbf{n}(\mathbf{x}_i), \mathbf{u} \rangle > 0$ for all i . Hence, since $\mathbf{x}_i \in S[\mathbf{u}]$, for each i there exists $t_i > 0$ such that $\mathbf{x}_i + t_i \mathbf{u} \in \partial\mathcal{O}$. For all i large enough that $\mathbf{x}_i \in \mathfrak{B}(\mathbf{x}, \tau/2)$, we necessarily have $t_i > \tau/2$. On the other hand, t_i is bounded above by the diameter of the object. Because t_i is bounded, it has a convergent subsequence $t_{i_1}, t_{i_2}, \dots \rightarrow \hat{t}$, with $\hat{t} > \tau/2$. Moreover, we have

$$\lim_{j \rightarrow \infty} \mathbf{x}_{i_j} + t_{i_j} \mathbf{u} = \mathbf{x} + \hat{t} \mathbf{u}.\tag{E.1}$$

Because \mathcal{O} is closed, this point is in \mathcal{O} . Because $\hat{t} > \tau/2 > 0$, this implies that $\mathbf{x} \notin S[\mathbf{u}]^c$, a contradiction. Hence, for every $\mathbf{x} \in S[\mathbf{u}]^c$, there exists $r_0 > 0$ such that $\mathfrak{B}(\mathbf{x}, r_0) \cap \partial\mathcal{O} \subseteq S[\mathbf{u}]^c$, and so $S[\mathbf{u}]^c$ is relatively open and $S[\mathbf{u}]$ relatively closed. \square

Proof of Lemma 5.2. We use the notation¹¹

$$B[\mathbf{u}] \doteq \{\mathbf{x} \in \Phi \mid \langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle \leq 0\}.\tag{E.2}$$

This set contains those points \mathbf{x} that are necessarily shadowed, because $\mathbf{n}(\mathbf{x})$ has nonpositive inner product with the light direction. Intuitively, if we ignore $B[\mathbf{u}]$, the remainder $S[\mathbf{u}] \setminus B[\mathbf{u}]$ should contain cast shadows, and for any point \mathbf{x} in this set, the shadow retraction $\mathbf{x}^{\mathbf{u}}$ should exist. Furthermore, if \mathbf{x} lies on the boundary of a cast shadow, its shadow retraction should lie in some *edge* of the object. We next state an intermediate lemma which makes this precise:

¹¹“B” can be taken to stand for “back”.

Lemma E.1. *For all \mathbf{u} , and all $\mathbf{x} \in \Phi \cap (\partial S[\mathbf{u}] \setminus B[\mathbf{u}])$, $\mathbf{x}^{\mathbf{u}}$ exists, and $\mathbf{x}^{\mathbf{u}} \in E$.*

Proof. Fix \mathbf{u} . Consider $\mathbf{x} \in \Phi \cap (\partial S[\mathbf{u}] \setminus B[\mathbf{u}])$. Since $\mathbf{x} \in \Phi$, $\mathbf{n}(\mathbf{x})$ is well-defined. Since $\mathbf{x} \notin B[\mathbf{u}]$, $\langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle > 0$. Since $\mathbf{x} \in S[\mathbf{u}]$, $\langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle \nu(\mathbf{x}, \mathbf{u}) = 0$. This implies that $\mathbf{x}^{\mathbf{u}}$ exists, $\mathbf{x}^{\mathbf{u}} \neq \mathbf{x}$, and $\mathbf{x}^{\mathbf{u}}$ is an element of $\partial \mathcal{O}$. If $\mathbf{x}^{\mathbf{u}} \in E$, we are done. For purpose of contradiction, suppose that $\mathbf{x}^{\mathbf{u}} \in \text{relint}(\Delta)$ for some face Δ . Thus, there exists $\varepsilon > 0$ such that $\mathfrak{B}(\mathbf{x}^{\mathbf{u}}, \varepsilon) \cap \text{aff}(\Delta) \subseteq \Delta$. Since $\mathbf{x} \in \Phi$, $\mathbf{x} \in \text{relint}(\Delta')$ for some face Δ' . Hence, there also exists $\varepsilon' > 0$ such that $\mathfrak{B}(\mathbf{x}, \varepsilon') \cap \partial \mathcal{O} \subseteq \Delta'$.

Choose $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^3$ such that

$$\text{aff}(\Delta) = \{\mathbf{x}^{\mathbf{u}} + \alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 \mid \alpha_1, \alpha_2 \in \mathbb{R}\}.$$

Similarly, choose orthonormal vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ such that

$$\text{aff}(\Delta') = \{\mathbf{x} + \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 \mid \beta_1, \beta_2 \in \mathbb{R}\}.$$

We claim that $\mathbf{u} \notin \text{span}(\mathbf{w}_1, \mathbf{w}_2)$. Indeed, if not, then for small δ_t , $\mathbf{x}^{\mathbf{u}} - \delta_t \mathbf{u} \in \Delta$. This would imply that $\mathbf{x} + (t^*(\mathbf{x}, \mathbf{u}) - \delta_t) \mathbf{u} \in \partial \mathcal{O}$, contradicting the minimality of $t^*(\mathbf{x}, \mathbf{u})$. So,

$$\text{rank}([\mathbf{w}_1 \mid \mathbf{w}_2 \mid \mathbf{u}]) = 3.$$

Consider a generic point $\mathbf{x}' = \mathbf{x} + [\mathbf{v}_1 \mid \mathbf{v}_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \in \text{aff}(\Delta')$. We find t such that $\mathbf{x}' + t\mathbf{u} \in \text{aff}(\Delta)$. This is possible iff the equation

$$\mathbf{x}^{\mathbf{u}} + \mathbf{W}\boldsymbol{\alpha} = \mathbf{x} + t\mathbf{u} + \mathbf{V}\boldsymbol{\beta}$$

(with $\mathbf{W} = [\mathbf{w}_1 \mid \mathbf{w}_2]$, $\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2]$) has a solution $(\boldsymbol{\alpha}, t)$. Rearranging, we obtain

$$[\mathbf{W} \mid \mathbf{u}] \begin{bmatrix} \boldsymbol{\alpha} \\ -t \end{bmatrix} = [\mathbf{V} \mid \mathbf{u}] \begin{bmatrix} \boldsymbol{\beta} \\ -t^*(\mathbf{x}, \mathbf{u}) \end{bmatrix}.$$

Since the matrix on the left has full rank three, we have

$$\begin{bmatrix} \boldsymbol{\alpha} \\ -t \end{bmatrix} = [\mathbf{W} \mid \mathbf{u}]^{-1} [\mathbf{V} \mid \mathbf{u}] \begin{bmatrix} \boldsymbol{\beta} \\ -t^*(\mathbf{x}, \mathbf{u}) \end{bmatrix}.$$

When $\boldsymbol{\beta} = \mathbf{0}$, the solution is $\boldsymbol{\alpha} = \mathbf{0}$, $t = t^*(\mathbf{x}, \mathbf{u}) > 0$. Hence, we can find $\varepsilon'' \in (0, \varepsilon')$ such that $\|\boldsymbol{\beta}\|_2 \leq \varepsilon''$ implies that (i) $t > 0$, (ii) $\mathbf{x} + \mathbf{V}\boldsymbol{\beta} \in \Delta'$, (iii) $\Delta \ni \mathbf{x}^{\mathbf{u}} + \mathbf{W}\boldsymbol{\alpha} = \mathbf{x} + t\mathbf{u}$. Because $\varepsilon'' < \varepsilon'$, every $\mathbf{x}' \in \mathfrak{B}(\mathbf{x}, \varepsilon'') \cap \partial \mathcal{O}$ lies in Δ' , and therefore has an expression $\mathbf{x}' = \mathbf{x} + \mathbf{V}\boldsymbol{\beta}$ with $\|\boldsymbol{\beta}\| < \varepsilon''$. Here, the fact that $\|\boldsymbol{\beta}\| \leq \varepsilon''$ follows because \mathbf{V} has orthonormal columns. By properties (i)-(iii) above, $\mathbf{x}'^{\mathbf{u}}$ exists, $\mathbf{x}' \in S[\mathbf{u}]$, and so $\mathfrak{B}(\mathbf{x}, \varepsilon'') \cap \partial \mathcal{O} \subseteq S[\mathbf{u}]$. This implies that $\mathbf{x} \notin \partial S[\mathbf{u}]$. Hence, if $\mathbf{x}^{\mathbf{u}} \notin E$, $\mathbf{x} \notin \partial S[\mathbf{u}]$, and the proof is complete. \square

Proof of Lemma 5.2. If $\mathbf{x} \in \partial S[\mathbf{u}] \cap \Phi$, then $\mathbf{x} \in \text{relint}(\Delta)$ for some face Δ . If $\mathbf{x} \in B[\mathbf{u}]$, then $\text{relint}(\Delta) \subseteq S[\mathbf{u}]$, and $\mathbf{x} \notin \partial S[\mathbf{u}]$. Hence, $\mathbf{x} \in B[\mathbf{u}]^c$. By Lemma E.1, $\mathbf{x}^{\mathbf{u}}$ exists, and is an element of E . \square

F Proof of Theorem 5.3

Proof. Our goal is to bound

$$\|\bar{\mathcal{D}}[\mathbf{u}] - \bar{\mathcal{D}}[\mathbf{u}']\|_{L^2}^2 = \int |\bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}) - \bar{\mathcal{D}}[\mathbf{u}'](\mathbf{x})|^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}). \quad (\text{F.1})$$

Initial manipulations. Notice that for $\mathbf{x} \in S[\mathbf{u}] \cap S[\mathbf{u}']$, $\bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}) = \bar{\mathcal{D}}[\mathbf{u}'](\mathbf{x}) = 0$. For $\mathbf{x} \in S[\mathbf{u}]^c \cap S[\mathbf{u}']^c$,

$$|\bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}) - \bar{\mathcal{D}}[\mathbf{u}'](\mathbf{x})| = \rho(\mathbf{x}) |\langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle - \langle \mathbf{n}(\mathbf{x}), \mathbf{u}' \rangle| \leq \rho(\mathbf{x}) \|\mathbf{u} - \mathbf{u}'\|_2. \quad (\text{F.2})$$

This bound also holds for $\mathbf{x} \in B[\mathbf{u}] \cup B[\mathbf{u}']$. Thus, setting

$$\Gamma = B[\mathbf{u}] \cup B[\mathbf{u}'] \cup (S[\mathbf{u}] \cap S[\mathbf{u}']) \cup (S[\mathbf{u}']^c \cap S[\mathbf{u}]^c), \quad (\text{F.3})$$

we obtain

$$\|\mathcal{P}_\Gamma (\bar{\mathcal{D}}[\mathbf{u}] - \bar{\mathcal{D}}[\mathbf{u}'])\|_{L^2}^2 \leq \rho_\star^2 \text{area}(\Gamma) \|\mathbf{u} - \mathbf{u}'\|_2^2. \quad (\text{F.4})$$

Note that

$$\partial\mathcal{O} = \Gamma \uplus \{S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])\} \uplus \{S[\mathbf{u}'] \setminus (S[\mathbf{u}] \cup B[\mathbf{u}'])\}, \quad (\text{F.5})$$

where \uplus denotes disjoint union. Consider $S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])$. Introduce a notation

$$\bar{\mathbf{u}}(r) = \frac{(1-r)\mathbf{u}' + r\mathbf{u}}{\|(1-r)\mathbf{u}' + r\mathbf{u}\|_2}, \quad r \in [0, 1], \quad (\text{F.6})$$

and set

$$r^\star(\mathbf{x}) = \inf \{r \in [0, 1] \mid \mathbf{x} \in S[\bar{\mathbf{u}}(r)]\}. \quad (\text{F.7})$$

For all $\mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])$, since $\mathbf{u} = \bar{\mathbf{u}}(1)$ and $\mathbf{x} \in S[\mathbf{u}]$, $r^\star(\mathbf{x}) \leq 1$ is finite. We have

$$\begin{aligned} |\bar{\mathcal{D}}[\mathbf{u}](\mathbf{x}) - \bar{\mathcal{D}}[\mathbf{u}'](\mathbf{x})| &= \bar{\mathcal{D}}[\mathbf{u}'](\mathbf{x}) = \rho(\mathbf{x}) \langle \mathbf{n}(\mathbf{x}), \mathbf{u}' \rangle \\ &= \rho(\mathbf{x}) \langle \mathbf{n}(\mathbf{x}), \mathbf{u}' - \bar{\mathbf{u}}(r^\star(\mathbf{x})) \rangle + \rho(\mathbf{x}) \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^\star(\mathbf{x})) \rangle \\ &\leq \rho(\mathbf{x}) \|\mathbf{u}' - \bar{\mathbf{u}}(r^\star(\mathbf{x}))\|_2 + \rho(\mathbf{x}) \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^\star(\mathbf{x})) \rangle. \end{aligned} \quad (\text{F.8})$$

So, we have

$$\begin{aligned} &\|\mathcal{P}_{S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])} (\bar{\mathcal{D}}[\mathbf{u}] - \bar{\mathcal{D}}[\mathbf{u}'])\|_{L^2}^2 \\ &\leq 2\rho_\star^2 \cdot \text{area}(S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])) \|\mathbf{u} - \mathbf{u}'\|_2^2 \\ &\quad + 2\rho_\star^2 \cdot \int_{\mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^\star(\mathbf{x})) \rangle^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) \end{aligned} \quad (\text{F.9})$$

and our task is to bound the final integral.¹² We will show the following:

(Δ) For all $\mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])$ such that $\mathbf{x} \in \Phi$, we have $\mathbf{x} \in \partial S[\bar{\mathbf{u}}(r^\star(\mathbf{x}))]$.

The intuition behind this claim is straightforward: as the light direction moves from \mathbf{u}' to \mathbf{u} , the first time that \mathbf{x} falls into shadow, it must lie in the boundary of the shadow (imagine the boundary of the shadow sweeping across the face Δ_j). Obtaining this as a rigorous consequence of our assumptions requires some manipulation, which we perform below.

¹²Under our assumptions, the function $\zeta(\mathbf{x}) = \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^\star(\mathbf{x})) \rangle^2$ can be shown to be piecewise rational, with pieces defined on semialgebraic sets. This implies that $\zeta(\mathbf{x})$ is measurable, and the integral in (F.9) is well-defined.

Proving (Δ) . Since $\mathbf{x} \in S[\mathbf{u}]$, and $\mathbf{u} = \bar{\mathbf{u}}(1)$, $r^*(\mathbf{x}) \leq 1$ is finite. Notice that $\bar{\mathbf{u}}(r)$ is continuous in r . Take $r_i \rightarrow r^*(\mathbf{x})$, with $\mathbf{x} \in S[\bar{\mathbf{u}}(r_i)]$. Then $\bar{\mathbf{u}}(r_i) \rightarrow \bar{\mathbf{u}}(r^*(\mathbf{x}))$. If $\mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])$ then $\langle \mathbf{n}(\mathbf{x}), \mathbf{u} \rangle > 0$, $\langle \mathbf{n}(\mathbf{x}), \mathbf{u}' \rangle > 0$, and so for any $r \in [0, 1]$, $\langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r) \rangle > 0$. Hence, for any r such that $\mathbf{x} \in S[\bar{\mathbf{u}}(r)]$, it must be that $\nu(\mathbf{x}, \mathbf{u}) = 0$, and $\mathbf{x}^{\bar{\mathbf{u}}(r)}$ exists. So, for each of our sequence of r_i , $\mathbf{x}^{\bar{\mathbf{u}}(r_i)} \in \partial\mathcal{O}$ exists:

$$\mathbf{x} + t^*(\mathbf{x}, \bar{\mathbf{u}}(r_i)) \bar{\mathbf{u}}(r_i) \in \partial\mathcal{O},$$

where we recall that $t^*(\mathbf{x}, \bar{\mathbf{u}}(r_i)) = \inf\{t > 0 \mid \mathbf{x} + t\bar{\mathbf{u}}(r_i) \in \partial\mathcal{O}\}$. Let Δ be the face containing \mathbf{x} . Since $\mathbf{x} \in \Phi$, $\mathbf{x} \in \text{relint}(\Delta)$, and there exists $\varepsilon > 0$ such that $\mathfrak{B}(\mathbf{x}, \varepsilon) \cap \partial\mathcal{O} \subseteq \Delta$. Since $\langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r_i) \rangle > 0$, if $t > 0$ is such that $\mathbf{x} + t\bar{\mathbf{u}}(r_i) \in \partial\mathcal{O}$, then $t > \varepsilon$. Hence, for every i , $t^*(\mathbf{x}, \bar{\mathbf{u}}(r_i)) \geq \varepsilon > 0$. Because \mathcal{O} is bounded, the $\beta_i \doteq t^*(\mathbf{x}, \bar{\mathbf{u}}(r_i))$ are bounded. Hence, the sequence (β_i) has a convergent subsequence β_{i_j} : $\lim_{j \rightarrow \infty} \beta_{i_j} = \beta_*$ for some β_* . From the previous discussion $\beta_* \geq \varepsilon$. Moreover, $\bar{\mathbf{u}}(r_{i_j}) \rightarrow \bar{\mathbf{u}}(r^*(\mathbf{x}))$. Hence

$$\mathbf{x} + \beta_{i_j} \bar{\mathbf{u}}(r_{i_j}) \rightarrow \mathbf{x} + \beta_* \bar{\mathbf{u}}(r^*(\mathbf{x})).$$

Because each element of the left hand side is in $\partial\mathcal{O}$, and $\partial\mathcal{O}$ is closed, the limit is in $\partial\mathcal{O}$. Because $\beta_* > 0$, the right hand side is not equal to \mathbf{x} . We conclude that $\mathbf{x} \in S[\bar{\mathbf{u}}(r^*(\mathbf{x}))]$.

It is left to show that \mathbf{x} lives in the relative boundary $\partial S[\bar{\mathbf{u}}(r^*(\mathbf{x}))]$ of this set. Choose $r' \in (0, r^*)$, and note that $\langle \bar{\mathbf{u}}(r'), \mathbf{n}(\mathbf{x}) \rangle > 0$. Notice that

$$\text{aff}(\Delta) = \{\mathbf{x}' \mid \langle \mathbf{n}(\mathbf{x}), \mathbf{x}' \rangle = \langle \mathbf{n}(\mathbf{x}), \mathbf{x} \rangle\}.$$

Hence, for any $\tau > 0$, if we set

$$s = \tau \frac{\langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle}{\langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r') \rangle},$$

then

$$\mathbf{x}' \doteq \mathbf{x} - \tau \bar{\mathbf{u}}(r^*(\mathbf{x})) + s \bar{\mathbf{u}}(r') \in \text{aff}(\Delta).$$

Suppose, for purpose of contradiction, that $\mathbf{x} \in \text{relint}(S[\bar{\mathbf{u}}(r^*(\mathbf{x}))])$. Then, there exists $\tau_0 > 0$ such that for $\tau \in (0, \tau_0)$, we have

$$\mathbf{x}' \in S[\bar{\mathbf{u}}(r^*(\mathbf{x}))].$$

Moreover, if $\mathbf{x}' \in \mathfrak{B}(\mathbf{x}, \varepsilon/2)$, and $t > 0$ is such that $\mathbf{x}' + t\bar{\mathbf{u}}(r^*(\mathbf{x})) \in \partial\mathcal{O}$, then $t \geq \varepsilon/2$. Choose $\tau > 0$ small enough that $\tau < \tau_0$, $\|\mathbf{x}' - \mathbf{x}\|_2 < \varepsilon/2$, and $\tau < \varepsilon/2$.

With these choices, $\mathbf{x}' \in S[\bar{\mathbf{u}}(r^*(\mathbf{x}))]$, and there exists $t \geq \varepsilon/2 > \tau$ such that

$$\mathbf{x}' + t\bar{\mathbf{u}}(r^*(\mathbf{x})) \in \partial\mathcal{O}.$$

Write

$$\mathbf{x}' + t\bar{\mathbf{u}}(r^*(\mathbf{x})) = \mathbf{x} + s\bar{\mathbf{u}}(r') + (t - \tau)\bar{\mathbf{u}}(r^*(\mathbf{x})) = \mathbf{x} + s'\mathbf{v},$$

with

$$\mathbf{v} = \frac{s\bar{\mathbf{u}}(r') + (t - \tau)\bar{\mathbf{u}}(r^*)}{\|s\bar{\mathbf{u}}(r') + (t - \tau)\bar{\mathbf{u}}(r^*)\|_2}$$

and

$$s' = \|s\bar{\mathbf{u}}(r') + (t - \tau)\bar{\mathbf{u}}(r^*)\|_2.$$

Since $\mathbf{v} = \bar{\mathbf{u}}(r'')$ for some $r'' > 0$ which is strictly smaller than $r^*(\mathbf{x})$, and $\mathbf{x} \in S[\bar{\mathbf{u}}(r'')]$, this contradicts the definition of r^* as the infimum. Hence, $\mathbf{x} \in \partial S[\bar{\mathbf{u}}(r^*(\mathbf{x}))]$.

Bounding the integral. *The main utility of (Δ) is that it allows us to organize our calculations in terms of the points that cast the shadows, rather than the points that are shadowed. We next reduce the problem of obtaining the desired bound to that of showing one key inequality, (F.12). We show how this inequality implies the desired result, and then return to show that this inequality indeed holds.*

To lighten the notation, we write

$$H = S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}]), \quad H_\Phi = H \cap \Phi. \quad (\text{F.10})$$

Since every $\mathbf{x} \in H_\Phi$ satisfies $\mathbf{x} \in \partial S[\bar{\mathbf{u}}(r^*(\mathbf{x}))]$, by the Lemma E.1, $\mathbf{x}^{\bar{\mathbf{u}}(r^*(\mathbf{x}))}$ exists, and is an element of some edge $e^*(\mathbf{x})$. For each $e \in \mathcal{E}$, let

$$\Xi_{e,\Delta} \doteq \{\mathbf{x} \in H_\Phi \cap \Delta \mid e^*(\mathbf{x}) = e\}. \quad (\text{F.11})$$

By the above discussion,

$$H_\Phi = \bigcup_{e,\Delta} \Xi_{e,\Delta}.$$

Below, we will demonstrate the following key inequality:¹³

$$\sum_{e,\Delta} \int_{\mathbf{x} \in \Xi_{e,\Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) \leq 8\sqrt{2} \cdot \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \chi_\star. \quad (\text{F.12})$$

The proof of this inequality will consist of several steps, which are carried out below. We first show that this inequality implies the desired result. Notice that by definition,

$$\begin{aligned} \int_{\mathbf{x} \in H} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) &= \int_{\mathbf{x} \in H_\Phi} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) \\ &\leq 8\sqrt{2} \cdot \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \chi_\star. \end{aligned} \quad (\text{F.13})$$

Combining with (F.9), we obtain

$$\begin{aligned} &\|\mathcal{P}_{S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])}(\mathcal{D}[\mathbf{u}] - \mathcal{D}[\mathbf{u}'])\|_{L^2}^2 \\ &\leq 2\rho_\star^2 \text{area}(S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])) \|\mathbf{u} - \mathbf{u}'\|_2^2 + 16\sqrt{2} \rho_\star^2 \text{diam}(\mathcal{O}) \chi_\star \|\mathbf{u} - \mathbf{u}'\|_2. \end{aligned} \quad (\text{F.14})$$

By symmetry, we also obtain

$$\begin{aligned} &\|\mathcal{P}_{S[\mathbf{u}'] \setminus (S[\mathbf{u}] \cup B[\mathbf{u}'])}(\mathcal{D}[\mathbf{u}] - \mathcal{D}[\mathbf{u}'])\|_{L^2}^2 \\ &\leq 2\rho_\star^2 \text{area}(S[\mathbf{u}'] \setminus (S[\mathbf{u}] \cup B[\mathbf{u}']))) \|\mathbf{u} - \mathbf{u}'\|_2^2 + 16\sqrt{2} \rho_\star^2 \text{diam}(\mathcal{O}) \chi_\star \|\mathbf{u} - \mathbf{u}'\|_2. \end{aligned} \quad (\text{F.15})$$

Combining these two bounds with (F.4) and (F.5), we obtain the claimed result. We are just left to verify (F.12).

Proof of Key Inequality (F.12). *Rather than directly proving (F.12), which requires us to bound integrals over $\Xi_{e,\Delta}$, we first demonstrate a bound over a much simpler set (which turns out to be a*

¹³Below, we will show that the $\Xi_{e,\Delta}$ are semialgebraic sets, and hence measurable. Thus, the integrals in (F.12) are well-defined.

quadrilateral $Q \subseteq \Delta$), and then show that we can arbitrarily well-approximate the domain of interest using finite collections of such quadrilaterals, to obtain the desired bound.

For each $\mathbf{z} \in e$, and each Δ , let

$$\tau_{e,\Delta}(\mathbf{z}) = \{r \in [0, 1] \mid \mathbf{z}_{\bar{\mathbf{u}}(r)} \text{ exists, and } \mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e,\Delta}\} \quad (\text{F.16})$$

It is immediate that $\Xi_{e,\Delta} = \{\mathbf{z}_{\bar{\mathbf{u}}(r)} \mid \mathbf{z} \in e, r \in \tau_{e,\Delta}(\mathbf{z})\}$. Let $[\mathbf{v}, \mathbf{w}] = \text{conv}\{\mathbf{v}, \mathbf{w}\}$. Simple geometric reasoning shows that if $[\mathbf{v}, \mathbf{w}] \subseteq e$, and $[r_1, r_2] \subseteq \tau_{e,\Delta}(\mathbf{z})$ for every $\mathbf{z} \in [\mathbf{v}, \mathbf{w}]$ (i.e., $[r_1, r_2] \subseteq \bigcap_{\mathbf{z} \in [\mathbf{v}, \mathbf{w}]} \tau_{e,\Delta}(\mathbf{z})$), the set

$$\mathcal{Q}([\mathbf{v}, \mathbf{w}], [r_1, r_2]) \doteq \{\mathbf{z}_{\bar{\mathbf{u}}(r)} \mid r \in [r_1, r_2], \mathbf{z} \in [\mathbf{v}, \mathbf{w}]\} \subset \Xi_{e,\Delta} \quad (\text{F.17})$$

is a quadrilateral. We will show the following:

(\square) Let $[\mathbf{v}, \mathbf{w}] \subseteq e$. Suppose that $[r_1, r_2] \subseteq \bigcap_{\mathbf{z} \in [\mathbf{v}, \mathbf{w}]} \tau_{e,\Delta}(\mathbf{z})$. Let

$$Q = \mathcal{Q}([\mathbf{v}, \mathbf{w}], [r_1, r_2]).$$

Then we have that

$$\int_{\mathbf{x} \in Q} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) \leq 8\sqrt{2} \cdot \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \|\mathbf{v} - \mathbf{w}\|_2 |r_2 - r_1|. \quad (\text{F.18})$$

We show the claim (\square). If $\mathbf{v} - \mathbf{w} \in \text{span}(\mathbf{u}, \mathbf{u}')$, then Q has measure zero, the integral on the left hand side is zero, and the bound holds trivially. Suppose that $\mathbf{v} - \mathbf{w} \notin \text{span}(\mathbf{u}, \mathbf{u}')$. Notice that if $\mathbf{x}_{\bar{\mathbf{u}}(r)} \in e$, then there exists a solution (s_1, s_2, s_3) to the system of equations $\mathbf{x} + s_1\mathbf{u} + s_2\mathbf{u}' = \mathbf{v}s_3 + \mathbf{w}(1 - s_3)$. When $\mathbf{v} - \mathbf{w} \notin \text{span}(\mathbf{u}, \mathbf{u}')$ this system has at most one solution, and so for each \mathbf{x} there is at most one r such that

$$\mathbf{x}_{\bar{\mathbf{u}}(r)} \in e.$$

Now, for $\mathbf{x} \in Q \subseteq \Xi_{e,\Delta}$, we have $\mathbf{x}_{\bar{\mathbf{u}}(r^*(\mathbf{x}))} \in e$. Moreover, by construction of Q , $\mathbf{x} = \mathbf{z}_{\bar{\mathbf{u}}(r)}$ for some $\mathbf{z} \in [\mathbf{v}, \mathbf{w}]$ and $r \in [r_1, r_2]$. Hence, it must be that $r = r^*(\mathbf{x})$, and so $r^*(\mathbf{x}) \in [r_1, r_2]$.

Set $\mathbf{u}_1 = \bar{\mathbf{u}}(r_1)$, $\mathbf{u}_2 = \bar{\mathbf{u}}(r_2)$. Notice that $\mathbf{n}(\mathbf{x})$ is constant over Q . We abbreviate it by \mathbf{n} . Suppose that $\langle \mathbf{n}, \bar{\mathbf{u}}(r) \rangle$ is maximized over $[r_1, r_2]$ at $r = r_1$. Then

$$\int_{\mathbf{x} \in Q} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial\mathcal{O}}(\mathbf{x}) \leq (\mathbf{n}^T \mathbf{u}_1)^2 \text{area}(Q). \quad (\text{F.19})$$

Let \mathbf{x}_0 be an arbitrary point in Δ . Then

$$\text{aff}(\Delta) = \{\mathbf{x} \mid \langle \mathbf{n}, \mathbf{x} \rangle = \langle \mathbf{n}, \mathbf{x}_0 \rangle\}.$$

Using this expression, we can write the shadow projection $\mathbf{z}_{\mathbf{u}}$ as

$$\mathbf{z}_{\mathbf{u}} = \left(\mathbf{I} - \frac{\mathbf{u}\mathbf{n}^T}{\mathbf{n}^T \mathbf{u}} \right) \mathbf{z} + \frac{\mathbf{n}^T \mathbf{x}_0}{\mathbf{n}^T \mathbf{u}} \mathbf{u} \quad (\text{F.20})$$

The set Q is a quadrilateral, with sides $[\mathbf{w}_{\mathbf{u}_1}, \mathbf{v}_{\mathbf{u}_1}]$, $[\mathbf{v}_{\mathbf{u}_1}, \mathbf{v}_{\mathbf{u}_2}]$, $[\mathbf{v}_{\mathbf{u}_2}, \mathbf{w}_{\mathbf{u}_2}]$, $[\mathbf{w}_{\mathbf{u}_1}, \mathbf{w}_{\mathbf{u}_2}]$. We can calculate

$$\begin{aligned} \mathbf{v}_{\mathbf{u}_1} - \mathbf{v}_{\mathbf{u}_2} &= \left(\mathbf{v} - \mathbf{u}_1 \frac{\mathbf{n}^T \mathbf{v}}{\mathbf{n}^T \mathbf{u}_1} + \frac{\mathbf{n}^T \mathbf{x}_0}{\mathbf{n}^T \mathbf{u}_1} \mathbf{u}_1 \right) - \left(\mathbf{v} - \mathbf{u}_2 \frac{\mathbf{n}^T \mathbf{v}}{\mathbf{n}^T \mathbf{u}_2} + \frac{\mathbf{n}^T \mathbf{x}_0}{\mathbf{n}^T \mathbf{u}_2} \mathbf{u}_2 \right) \\ &= \left(\frac{\mathbf{u}_1}{\mathbf{n}^T \mathbf{u}_1} - \frac{\mathbf{u}_2}{\mathbf{n}^T \mathbf{u}_2} \right) (\mathbf{n}^T (\mathbf{x}_0 - \mathbf{v})), \end{aligned}$$

and similarly, $\mathbf{w}_{\mathbf{u}_1} - \mathbf{w}_{\mathbf{u}_2} = \left(\frac{\mathbf{u}_1}{\mathbf{n}^T \mathbf{u}_1} - \frac{\mathbf{u}_2}{\mathbf{n}^T \mathbf{u}_2} \right) (\mathbf{n}^T (\mathbf{x}_0 - \mathbf{w}))$. Since these differences are scalar multiples of the common vector $\frac{\mathbf{u}_1}{\mathbf{n}^T \mathbf{u}_1} - \frac{\mathbf{u}_2}{\mathbf{n}^T \mathbf{u}_2}$, the two sides $[\mathbf{v}_{\mathbf{u}_1}, \mathbf{v}_{\mathbf{u}_2}]$ and $[\mathbf{w}_{\mathbf{u}_1}, \mathbf{w}_{\mathbf{u}_2}]$ are parallel.

Let ℓ^\perp denote the orthogonal length

$$\ell^\perp \doteq \left\| \mathbf{P}_{(\mathbf{v}_{\mathbf{u}_1} - \mathbf{v}_{\mathbf{u}_2})^\perp} (\mathbf{w}_{\mathbf{u}_1} - \mathbf{v}_{\mathbf{u}_1}) \right\|_2. \quad (\text{F.21})$$

We have

$$\begin{aligned} \text{area}(Q) &= \frac{\|\mathbf{v}_{\mathbf{u}_1} - \mathbf{v}_{\mathbf{u}_2}\|_2 + \|\mathbf{w}_{\mathbf{u}_1} - \mathbf{w}_{\mathbf{u}_2}\|_2}{2} \times \ell^\perp \\ &= \frac{|\mathbf{n}^T (\mathbf{v} - \mathbf{x}_0)| + |\mathbf{n}^T (\mathbf{w} - \mathbf{x}_0)|}{2} \left\| \frac{\mathbf{u}_1}{\mathbf{n}^T \mathbf{u}_1} - \frac{\mathbf{u}_2}{\mathbf{n}^T \mathbf{u}_2} \right\|_2 \times \ell^\perp \end{aligned} \quad (\text{F.22})$$

Since $\mathbf{v} \in \mathcal{O}$, and $\mathbf{v}_{\mathbf{u}_2} \in \Delta \subseteq \mathcal{O}$, we have $\|\mathbf{v} - \mathbf{v}_{\mathbf{u}_2}\|_2 \leq \text{diam}(\mathcal{O})$. If we consider the right triangle formed by \mathbf{v} , $\mathbf{v}_{\mathbf{u}_2}$, and $\mathbf{v}_{\text{proj}} = \mathbf{v} - \mathbf{n} \mathbf{n}^T (\mathbf{v} - \mathbf{x}_0)$ (the orthogonal projection of \mathbf{v} onto $\text{aff}(\Delta)$), we have

$$\begin{aligned} |\mathbf{n}^T (\mathbf{v} - \mathbf{x}_0)| &= \|\mathbf{v} - (\mathbf{v} - \mathbf{n} \mathbf{n}^T (\mathbf{v} - \mathbf{x}_0))\|_2, \\ &= \|\mathbf{v} - \mathbf{v}_{\mathbf{u}_2}\|_2 \cdot \sin \angle (\mathbf{v} - \mathbf{v}_{\mathbf{u}_2}, \mathbf{v}_{\text{proj}} - \mathbf{v}_{\mathbf{u}_2}), \\ &= \|\mathbf{v} - \mathbf{v}_{\mathbf{u}_2}\|_2 \times \mathbf{n}^T \mathbf{u}_2, \\ &\leq \text{diam}(\mathcal{O}) \times \mathbf{n}^T \mathbf{u}_2. \end{aligned} \quad (\text{F.23})$$

A similar inequality holds for $|\mathbf{n}^T (\mathbf{w} - \mathbf{x}_0)|$. Together, this implies that

$$\text{area}(Q) \leq \text{diam}(\mathcal{O}) \times \mathbf{n}^T \mathbf{u}_2 \times \left\| \frac{\mathbf{u}_1}{\mathbf{n}^T \mathbf{u}_1} - \frac{\mathbf{u}_2}{\mathbf{n}^T \mathbf{u}_2} \right\|_2 \times \ell^\perp. \quad (\text{F.24})$$

Thus, we have

$$\int_{\mathbf{x} \in Q} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \leq \mathbf{n}^T \mathbf{u}_1 \times \|\mathbf{n}^T \mathbf{u}_2 \mathbf{u}_1 - \mathbf{n}^T \mathbf{u}_1 \mathbf{u}_2\|_2 \times \text{diam}(\mathcal{O}) \times \ell^\perp. \quad (\text{F.25})$$

Using the triangle inequality, it is easy to show that

$$\|\mathbf{n}^T \mathbf{u}_2 \mathbf{u}_1 - \mathbf{n}^T \mathbf{u}_1 \mathbf{u}_2\|_2 \leq 2 \|\mathbf{u}_1 - \mathbf{u}_2\|_2. \quad (\text{F.26})$$

Using the general fact that for nonzero vectors \mathbf{a}, \mathbf{b} ,

$$\left\| \frac{\mathbf{a}}{\|\mathbf{a}\|_2} - \frac{\mathbf{b}}{\|\mathbf{b}\|_2} \right\|_2 \leq \frac{2 \|\mathbf{a} - \mathbf{b}\|_2}{\max\{\|\mathbf{a}\|_2, \|\mathbf{b}\|_2\}}, \quad (\text{F.27})$$

and the fact that when $\|\mathbf{u} - \mathbf{u}'\|_2 \leq \sqrt{2}$, $\|r\mathbf{u} + (1-r)\mathbf{u}'\|_2 \geq 1/\sqrt{2}$ for all r in $[0, 1]$, we have

$$\begin{aligned} \|\mathbf{u}_1 - \mathbf{u}_2\|_2 &\leq \frac{2 \|r_1 \mathbf{u} + (1-r_1) \mathbf{u}' - (r_2 \mathbf{u} + (1-r_2) \mathbf{u}')\|_2}{\|r_1 \mathbf{u} + (1-r_1) \mathbf{u}'\|_2} \\ &\leq 2\sqrt{2} \cdot \|\mathbf{u} - \mathbf{u}'\|_2 |r_1 - r_2|, \end{aligned} \quad (\text{F.28})$$

Putting together (F.25), (F.26) and (F.28), we get

$$\int_{\mathbf{x} \in Q} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \leq 4\sqrt{2} \times \mathbf{n}^T \mathbf{u}_1 \times \ell^\perp \times \text{diam}(\mathcal{O}) \times \|\mathbf{u} - \mathbf{u}'\|_2 \times |r_1 - r_2|.$$

Finally, using the expression for $\mathbf{v}_{\mathbf{u}_1}$ and $\mathbf{w}_{\mathbf{u}_1}$, we obtain

$$\begin{aligned} \mathbf{n}^T \mathbf{u}_1 \times \ell^\perp &\leq \mathbf{n}^T \mathbf{u}_1 \|\mathbf{v}_{\mathbf{u}_1} - \mathbf{w}_{\mathbf{u}_1}\|_2 \\ &= \|(\mathbf{n}^T \mathbf{u}_1)(\mathbf{v} - \mathbf{w}) - \mathbf{u}_1 \mathbf{n}^T(\mathbf{v} - \mathbf{w})\|_2 \\ &\leq 2 \|\mathbf{v} - \mathbf{w}\|_2. \end{aligned} \tag{F.29}$$

This completes the proof of (\square) for the case when $\langle \mathbf{n}, \bar{\mathbf{u}}(r) \rangle$ is maximized at $r = r_1$. If $\langle \mathbf{n}, \bar{\mathbf{u}}(r) \rangle$ is instead maximized at $r = r_2$, we may simply repeat the above argument, interchanging \mathbf{u}_1 and \mathbf{u}_2 . If $\langle \mathbf{n}, \bar{\mathbf{u}}(r) \rangle$ is instead maximized at some $r_0 \in (r_1, r_2)$, we may partition Q into two sub-quadrilaterals, indexed by $[r_0, r_1]$ and $[r_1, r_2]$, respectively, and then apply the argument to each. This establishes (\square) .

Our approach, then, is to discretize the domain of integration and apply (\square) . We make the following technical claim regarding approximation of the domain of integration $\Xi_{e,\Delta}$ by quadrilaterals:

(\diamond) For each edge e and face Δ , and any $\varepsilon > 0$, there exists a finite collection of segments

$$[\mathbf{a}_1, \mathbf{b}_1] \cup [\mathbf{a}_2, \mathbf{b}_2] \cup \dots \cup [\mathbf{a}_N, \mathbf{b}_N] \subseteq e$$

with disjoint relative interiors, and a collection of interior-disjoint intervals

$$\left([r_{2j-1}^{(1)}, r_{2j}^{(1)}] \right)_{j=1}^{n_1}, \dots, \left([r_{2j-1}^{(N)}, r_{2j}^{(N)}] \right)_{j=1}^{n_N}$$

with the following properties:

$$(i) \quad [r_{2j-1}^{(i)}, r_{2j}^{(i)}] \subseteq \bigcap_{\mathbf{z} \in [\mathbf{a}_i, \mathbf{b}_i]} \tau_{e,\Delta}(\mathbf{z}), \tag{F.30}$$

which implies that $Q_{ij} \doteq \mathcal{Q}([\mathbf{a}_i, \mathbf{b}_i], [r_{2j-1}^{(i)}, r_{2j}^{(i)}]) \subseteq \Xi_{e,\Delta}$, and

$$(ii) \quad \mu_{\partial \mathcal{O}} \left(\Xi_{e,\Delta} \setminus \bigcup_{i,j} Q_{ij} \right) \leq \varepsilon. \tag{F.31}$$

We will show (\diamond) below. Let us first examine its implications. We have

$$\begin{aligned} &\int_{\mathbf{x} \in \Xi_{e,\Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \\ &\leq \left(\sum_{i,j} \int_{\mathbf{x} \in Q_{ij}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \right) + \mu \left(\Xi_{e,\Delta} \setminus \bigcup_{i,j} Q_{ij} \right) \sup_{\mathbf{x} \in \partial \mathcal{O}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2, \\ &\leq \sum_{i,j} 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \|\mathbf{b}_i - \mathbf{a}_i\|_2 |r_{2j}^{(i)} - r_{2j-1}^{(i)}| + \varepsilon, \end{aligned}$$

where the first term follows from (\square) .

Consider the product $e \times [0, 1]$. The rectangles $[\mathbf{a}_i, \mathbf{b}_i] \times [r_{2j-1}^{(i)}, r_{2j}^{(i)}]$ have disjoint interiors. So,

$$\sum_{ij} \|\mathbf{b}_i - \mathbf{a}_i\|_2 |r_{2j}^{(i)} - r_{2j-1}^{(i)}| = \int_{\mathbf{z} \in e, r \in [0, 1]} \mathbb{1}_{(\mathbf{z}, r) \in \bigcup_{ij} \text{int}([\mathbf{a}_i, \mathbf{b}_i] \times [r_{2j-1}^{(i)}, r_{2j}^{(i)}])} d\mathbf{z} dr. \quad (\text{F.32})$$

By construction, for any pair $(\mathbf{z}, r) \in [\mathbf{a}_i, \mathbf{b}_i] \times [r_{2j-1}^{(i)}, r_{2j}^{(i)}]$ we have $\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}$. So,

$$\mathbb{1}_{(\mathbf{z}, r) \in \bigcup_{ij} \text{int}([\mathbf{a}_i, \mathbf{b}_i] \times [r_{2j-1}^{(i)}, r_{2j}^{(i)}])} \leq \mathbb{1}_{\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}},$$

and

$$\int_{\mathbf{x} \in \Xi_{e, \Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \int_{\mathbf{z} \in e, r \in [0, 1]} \mathbb{1}_{\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}} d\mathbf{z} dr + \varepsilon.$$

Since this holds for every $\varepsilon > 0$, we have

$$\int_{\mathbf{x} \in \Xi_{e, \Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \int_{\mathbf{z} \in e, r \in [0, 1]} \mathbb{1}_{\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}} d\mathbf{z} dr. \quad (\text{F.33})$$

Summing over e, Δ , we obtain

$$\sum_{e, \Delta} \int_{\mathbf{x} \in \Xi_{e, \Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \int_{r \in [0, 1]} \left(\sum_e \int_{\mathbf{z} \in e} \sum_{\Delta} \mathbb{1}_{\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}} d\mathbf{z} \right) dr.$$

Notice that for a given edge $e = [\mathbf{a}, \mathbf{b}]$, if it happens that $\mathbf{b} - \mathbf{a} \in \text{span}\{\mathbf{u}, \mathbf{u}'\}$, $\Xi_{e, \Delta}$ has measure zero. So, letting \mathcal{E}' denote the set of edges $[\mathbf{a}, \mathbf{b}]$ for which $\mathbf{b} - \mathbf{a} \notin \text{span}\{\mathbf{u}, \mathbf{u}'\}$, we have

$$\sum_{e, \Delta} \int_{\mathbf{x} \in \Xi_{e, \Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) \leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \int_{r \in [0, 1]} \left(\sum_{e \in \mathcal{E}'} \int_{\mathbf{z} \in e} \sum_{\Delta} \mathbb{1}_{\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}} d\mathbf{z} \right) dr.$$

It is not difficult to show that if $e \in \mathcal{E}'$, for each $\mathbf{x} \notin e$ there is at most one r such that $\mathbf{x}_{\bar{\mathbf{u}}(r)} \in e$. So, if $\mathbf{x} = \mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}$, it must be that $r = r^*(\mathbf{x})$. This implies (via (Δ)) that $\mathbf{x} \in \partial S[\bar{\mathbf{u}}(r)]$. Since $\mathbf{x} \in \Phi$ as well, and $\mathbf{z} = \mathbf{x}_{\bar{\mathbf{u}}(r)}$, we have $\mathbf{z} \in \chi[\bar{\mathbf{u}}(r)]$. So, for $e \in \mathcal{E}'$, $\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}$ implies that $\mathbf{z} \in \chi[\bar{\mathbf{u}}(r)]$. This, together with the fact that the sets $\Xi_{e, \Delta}$ and $\Xi_{e', \Delta'}$ are disjoint whenever $\Delta \neq \Delta'$ gives that

$$\sum_{\Delta} \mathbb{1}_{\mathbf{z}_{\bar{\mathbf{u}}(r)} \in \Xi_{e, \Delta}} \leq \mathbb{1}_{\mathbf{z} \in \chi[\bar{\mathbf{u}}(r)]},$$

and

$$\begin{aligned} \sum_{e, \Delta} \int_{\mathbf{x} \in \Xi_{e, \Delta}} \langle \mathbf{n}(\mathbf{x}), \bar{\mathbf{u}}(r^*(\mathbf{x})) \rangle^2 d\mu_{\partial \mathcal{O}}(\mathbf{x}) &\leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \int_{r \in [0, 1]} \left(\sum_{e \in \mathcal{E}'} \int_{\mathbf{z} \in e} \mathbb{1}_{\mathbf{z} \in \chi[\bar{\mathbf{u}}(r)]} d\mathbf{z} \right) dr \\ &\leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \int_{r \in [0, 1]} \text{length}(\chi[\bar{\mathbf{u}}(r)]) dr \\ &\leq 8\sqrt{2} \text{diam}(\mathcal{O}) \|\mathbf{u} - \mathbf{u}'\|_2 \chi_*, \end{aligned} \quad (\text{F.34})$$

as desired. To finish the proof, we are just left to show (\diamond) .

Demonstrating (\diamond) . We begin with the definition of $\Xi_{e,\Delta}$:

$$\Xi_{e,\Delta} = \left\{ \mathbf{x} \left| \begin{array}{l} \mathbf{x} \in \text{relint}(\Delta) \\ \mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}]) \\ \mathbf{x}^{\tilde{\mathbf{u}}(r^*(\mathbf{x}))} \in e \end{array} \right. \right\}. \quad (\text{F.35})$$

If the outward normal \mathbf{n} to Δ satisfies $\mathbf{n}^T \mathbf{u} \leq 0$, $\text{relint}(\Delta) \subseteq B[\mathbf{u}]$, and $\Xi_{e,\Delta}$ is empty, implying that (\diamond) is trivially satisfied. Similarly, if $\mathbf{n}^T \mathbf{u}' \leq 0$, $\Delta \subseteq S[\mathbf{u}']$, $\Xi_{e,\Delta}$ is empty, and (\diamond) is trivially satisfied. To fix notation, let $e = [\mathbf{a}, \mathbf{b}]$. If $\mathbf{b} - \mathbf{a} \in \text{span}\{\mathbf{u}, \mathbf{u}'\}$, then $\Xi_{e,\Delta}$ has measure zero, and (\diamond) is again trivially satisfied.

It remains to consider the case when $\mathbf{n}^T \mathbf{u} > 0$ and $\mathbf{n}^T \mathbf{u}' > 0$, and $\mathbf{b} - \mathbf{a} \notin \text{span}\{\mathbf{u}, \mathbf{u}'\}$. We will find it slightly more convenient to work with an unnormalized version of $\tilde{\mathbf{u}}$, by setting

$$\tilde{\mathbf{u}}(r) = r\mathbf{u} + (1-r)\mathbf{u}'. \quad (\text{F.36})$$

It is easy to check that $\mathbf{x}^{\tilde{\mathbf{u}}(r)}$ is defined if and only if $\mathbf{x}^{\tilde{\mathbf{u}}(r)}$ is defined, and $\mathbf{x}^{\tilde{\mathbf{u}}(r)} = \mathbf{x}^{\tilde{\mathbf{u}}(r)}$. So, we can rephrase our expression for $\Xi_{e,\Delta}$ as

$$\Xi_{e,\Delta} = \left\{ \mathbf{x} \left| \begin{array}{l} \mathbf{x} \in \text{relint}(\Delta) \\ \mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}]) \\ \mathbf{x}^{\tilde{\mathbf{u}}(r^*(\mathbf{x}))} \in e \end{array} \right. \right\}. \quad (\text{F.37})$$

To show that $\Xi_{e,\Delta}$ can be well-approximated by quadrilaterals of the desired form, it will be useful to work in coordinates. Let

$$\Delta = \text{conv}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}.$$

We can parameterize $\text{aff}(\Delta)$ in terms of $\mathbf{w} \in \mathbb{R}^2$ via

$$\mathbf{x}(\mathbf{w}) = \mathbf{v}_1 w_1 + \mathbf{v}_2 w_2 + \mathbf{v}_3(1 - w_1 - w_2) = \mathbf{V}\mathbf{w} + \mathbf{v}_3, \quad (\text{F.38})$$

with $\mathbf{V} = [\mathbf{v}_1 - \mathbf{v}_3 \mid \mathbf{v}_2 - \mathbf{v}_3] \in \mathbb{R}^{3 \times 2}$. Then

$$\Delta = \{\mathbf{x}(\mathbf{w}) \mid w_1 \geq 0, w_2 \geq 0, w_1 + w_2 \leq 1\}. \quad (\text{F.39})$$

Similarly, parameterize e via

$$\mathbf{z}(s) = s\mathbf{a} + (1-s)\mathbf{b}. \quad (\text{F.40})$$

Let

$$W_{e,\Delta} \doteq \{\mathbf{w} \mid \mathbf{x}(\mathbf{w}) \in \Xi_{e,\Delta}\} \subset \mathbb{R}^2. \quad (\text{F.41})$$

We will show that $W_{e,\Delta}$ is a *semialgebraic set* [Cos00]. As we will see, semialgebraic sets are sufficiently well-behaved to admit the approximation promised by (\diamond) . To show that $W_{e,\Delta}$ is semialgebraic, we take the conditions in (F.37) one at a time. First, notice that

$$\mathbf{x}(\mathbf{w}) \in \text{relint}(\Delta) \iff w_1 > 0, w_2 > 0, \text{ and } w_1 + w_2 < 1. \quad (\text{F.42})$$

Set

$$W_1 = \{\mathbf{w} \in \mathbb{R}^2 \mid w_1 > 0, w_2 > 0, \text{ and } w_1 + w_2 < 1\}. \quad (\text{F.43})$$

The set W_1 is semialgebraic.

Now, take the second condition in (F.37): $\mathbf{x} \in S[\mathbf{u}] \setminus (S[\mathbf{u}'] \cup B[\mathbf{u}])$. Since $\mathbf{n}^T \mathbf{u} > 0$, this condition reduces to $\mathbf{x} \in S[\mathbf{u}] \setminus S[\mathbf{u}']$. Moreover, since $\mathbf{n}^T \mathbf{u}' > 0$, this condition further reduces to

$$\exists t > 0 \text{ s.t. } \mathbf{x} + t\mathbf{u} \in \partial\mathcal{O}, \quad \text{and} \quad \nexists t' > 0 \text{ s.t. } \mathbf{x} + t'\mathbf{u}' \in \partial\mathcal{O}. \quad (\text{F.44})$$

For each Δ' , define a coordinate map $\mathbf{x}_{\Delta'}(\mathbf{w}')$ in the same manner as $\mathbf{x}(\mathbf{w})$. For each Δ' , write

$$S_{2,\Delta'} = \left\{ \mathbf{w}, \mathbf{w}', t \left| \begin{array}{l} \mathbf{x}(\mathbf{w}) + t\mathbf{u} = \mathbf{x}_{\Delta'}(\mathbf{w}') \\ t > 0 \\ w'_1 \geq 0, w'_2 \geq 0, w'_1 + w'_2 \leq 1 \end{array} \right. \right\} \subset \mathbb{R}^5. \quad (\text{F.45})$$

Write $\mathcal{P}_{\mathbf{w}}$ for the projection onto the \mathbf{w} coordinates, and

$$W_{2,\Delta'} = \mathcal{P}_{\mathbf{w}} S_{2,\Delta'}. \quad (\text{F.46})$$

With this definition, notice that $\exists t > 0$ such that $\mathbf{x}(\mathbf{w}) + t\mathbf{u} \in \partial\mathcal{O}$ if and only if $\mathbf{w} \in \bigcup_{\Delta' \neq \Delta} W_{2,\Delta'}$. Moreover, since each $S_{2,\Delta'}$ is defined by finitely many polynomial inequalities, each $S_{2,\Delta'}$ is semialgebraic. By the Tarski-Seidenberg theorem, each $W_{2,\Delta'}$ is also semialgebraic.

In a similar manner, define

$$S_{3,\Delta'} = \left\{ \mathbf{w}, \mathbf{w}', t' \left| \begin{array}{l} \mathbf{x}(\mathbf{w}) + t'\mathbf{u}' = \mathbf{x}_{\Delta'}(\mathbf{w}') \\ t' > 0 \\ w'_1 \geq 0, w'_2 \geq 0, w'_1 + w'_2 \leq 1 \end{array} \right. \right\} \quad (\text{F.47})$$

and $W_{3,\Delta'} = \mathcal{P}_{\mathbf{w}} S_{3,\Delta'}$. The $W_{3,\Delta'}$ are also semialgebraic. Combining these sets, we have that $\mathbf{x}(\mathbf{w}) \in S[\mathbf{u}] \setminus S[\mathbf{u}']$ if and only if

$$\mathbf{w} \in \left(\bigcup_{\Delta'} W_{2,\Delta'} \right) \setminus \left(\bigcup_{\Delta'} W_{3,\Delta'} \right) \doteq W_4. \quad (\text{F.48})$$

The set W_4 is produced from semialgebraic sets via finitely many set operations, and hence is semialgebraic.

The final condition in (F.37) that we need to consider is that $\mathbf{x}^{\tilde{\mathbf{u}}(r^*(\mathbf{x}))} \in e$. Because $\mathbf{b} - \mathbf{a} \notin \text{span}\{\mathbf{u}, \mathbf{u}'\}$, for each \mathbf{x} there exists at most one pair (\hat{t}, \hat{r}) with $\hat{t} > 0, \hat{r} \in [0, 1]$ such that $\mathbf{x} + \hat{t}\tilde{\mathbf{u}}(\hat{r}) \in e$. Hence, there exists at most one $\hat{r} \in [0, 1]$ such that $\mathbf{x}^{\tilde{\mathbf{u}}(\hat{r})} \in e$. For any given \hat{r} , $\mathbf{x}^{\tilde{\mathbf{u}}(\hat{r})} \in e$ if and only if the following two conditions are satisfied:

$$\exists \hat{t} \quad \text{s.t.} \quad \hat{t} > 0, \mathbf{x} + \hat{t}\tilde{\mathbf{u}}(\hat{r}) \in e \quad (\text{F.49})$$

$$\nexists (\hat{t}, \hat{t}') \quad \text{s.t.} \quad 0 < \hat{t}' < \hat{t}, \mathbf{x} + \hat{t}\tilde{\mathbf{u}}(\hat{r}) \in e, \mathbf{x} + \hat{t}'\tilde{\mathbf{u}}(\hat{r}) \in \partial\mathcal{O}. \quad (\text{F.50})$$

The first condition ensures that the ray $\mathbf{x} + \mathbb{R}_{++}\tilde{\mathbf{u}}(\hat{r})$ intersects e , while the second ensures that no other point of $\partial\mathcal{O}$ lies between \mathbf{x} and this intersection on the ray.

Set

$$S_5 = \left\{ \hat{r}, \hat{t}, \mathbf{w}, s \left| \begin{array}{l} \mathbf{x}(\mathbf{w}) + \hat{t}\tilde{\mathbf{u}}(\hat{r}) = \mathbf{z}(s) \\ s \in [0, 1] \\ \hat{t} > 0 \\ \hat{r} \in [0, 1] \end{array} \right. \right\} \quad (\text{F.51})$$

and $W_5 = \mathcal{P}_{\mathbf{w}} S_5$. Write

$$S_{6,\Delta'} = \left\{ \hat{r}, \hat{t}, \mathbf{w}, s, t', \mathbf{w}' \left| \begin{array}{l} \mathbf{x}(\mathbf{w}) + \hat{t}\tilde{\mathbf{u}}(\hat{r}) = \mathbf{z}(s) \\ s \in [0, 1] \\ \hat{t} > 0 \\ \hat{r} \in [0, 1] \\ 0 < t' < \hat{t} \\ w'_1 \geq 0, w'_2 \geq 0, w'_1 + w'_2 \leq 1 \\ \mathbf{x}(\mathbf{w}) + t'\tilde{\mathbf{u}}(\hat{r}) = \mathbf{x}_{\Delta'}(\mathbf{w}') \end{array} \right. \right\}, \quad (\text{F.52})$$

and $W_{6,\Delta'} = \mathcal{P}_{\mathbf{w}} S_{6,\Delta'}$. Then there exists $\hat{r} \in [0, 1]$ such that $\mathbf{x}(\mathbf{w})^{\tilde{\mathbf{u}}(\hat{r})} \in e$ if and only if

$$\mathbf{w} \in W_5 \setminus \bigcup_{\Delta' \neq \Delta} W_{6,\Delta'} \doteq W_7. \quad (\text{F.53})$$

Again, the set W_7 is semialgebraic.

Consider $\mathbf{w} \in W_7$. By construction this means that there exists \hat{r} such that $\mathbf{x}(\mathbf{w})^{\tilde{\mathbf{u}}(\hat{r})} \in e$. Moreover, by the above reasoning, this \hat{r} is the only r with this property. Is $\hat{r} = r^*(\mathbf{x}(\mathbf{w}))$? This is true if and only if there does not exist $r' \in (0, \hat{r})$ and $t' > 0$ with $\mathbf{x}(\mathbf{w}) + t'\tilde{\mathbf{u}}(r') \in \partial\mathcal{O}$. Let

$$S_{8,\Delta'} = \left\{ \hat{r}, \hat{t}, \mathbf{w}, s, r', t', \mathbf{w}' \left| \begin{array}{l} \mathbf{x}(\mathbf{w}) + \hat{t}\tilde{\mathbf{u}}(\hat{r}) = \mathbf{z}(s) \\ s \in [0, 1] \\ \hat{t} > 0 \\ \hat{r} \in [0, 1] \\ t' > 0 \\ 0 < r' < \hat{r} \\ w'_1 \geq 0, w'_2 \geq 0, w'_1 + w'_2 \leq 1 \\ \mathbf{x}(\mathbf{w}) + t'\tilde{\mathbf{u}}(\hat{r}) = \mathbf{x}_{\Delta'}(\mathbf{w}') \end{array} \right. \right\} \quad (\text{F.54})$$

and again set $W_{8,\Delta'} = \mathcal{P}_{\mathbf{w}} S_{8,\Delta'}$. Then we have

$$\mathbf{x}(\mathbf{w})^{\tilde{\mathbf{u}}(r^*(\mathbf{x}(\mathbf{w})))} \in e \iff \mathbf{w} \in W_7 \setminus \bigcup_{\Delta' \neq \Delta} W_{8,\Delta'} \doteq W_9. \quad (\text{F.55})$$

Hence, setting $W_{e,\Delta} = W_1 \cap W_4 \cap W_9$, we have

$$\mathbf{x}(\mathbf{w}) \in \Xi_{e,\Delta} \iff \mathbf{w} \in W_{e,\Delta}, \quad (\text{F.56})$$

and the set $W_{e,\Delta}$ is semialgebraic.

Our next task is to rewrite $W_{e,\Delta}$ in terms of the parameters s, r . Since $\mathbf{n}^T \mathbf{u} > 0$ and $\mathbf{n}^T \mathbf{u}' > 0$, for every $r \in [0, 1]$, $\mathbf{n}^T \tilde{\mathbf{u}}(r) > 0$. Hence, for every r, s there is a unique t such that

$$\mathbf{z}(s) - t\tilde{\mathbf{u}}(r) \in \text{aff}(\Delta).$$

In particular, there exists a unique $(\mathbf{w}, t) \in \mathbb{R}^2 \times \mathbb{R}$ such that

$$\mathbf{z}(s) - t\tilde{\mathbf{u}}(r) = \mathbf{x}(\mathbf{w}).$$

This gives a system of equations

$$[\mathbf{V} \mid \tilde{\mathbf{u}}(r)] \begin{bmatrix} \mathbf{w} \\ t \end{bmatrix} = \mathbf{z}(s) - \mathbf{v}_3. \quad (\text{F.57})$$

Under our assumptions, it is not difficult to show that there exists $\zeta > 0$ such that

$$\sigma_{\min}([\mathbf{V} \mid \tilde{\mathbf{u}}(r)]) \geq \zeta \quad (\text{F.58})$$

for all $r \in [0, 1]$.¹⁴ Hence, we can write the coordinates \mathbf{w} in this unique pair explicitly as a function of (r, s) :

$$\mathbf{w} = \Upsilon(r, s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} [\mathbf{V} \mid \tilde{\mathbf{u}}(r)]^{-1} (\mathbf{z}(s) - \mathbf{v}_3).$$

The components of Υ are rational functions of (r, s) , the denominator of which does not vanish. This implies that Υ is a *semialgebraic map* [Cos00]. Set

$$A_{e,\Delta} = \Upsilon^{-1}[W_{e,\Delta}] \subseteq [0, 1]^2. \quad (\text{F.59})$$

Since Υ is a semialgebraic map, $A_{e,\Delta}$ is also semialgebraic. $A_{e,\Delta} \subseteq [0, 1]^2$ is also bounded. Because semialgebraic sets are finite unions of intersections of sublevel sets of finitely many polynomial (and hence continuous) functions, bounded semialgebraic sets are Jordan measurable. This implies that for every $\eta > 0$, there exists an $M \in \mathbb{Z}$ and a collection of M interior disjoint rectangles $R_i = [r_{i,1}, r_{i,2}] \times [s_{i,1}, s_{i,2}]$ ($i = 1 \dots m$) such that

$$\mu \left(A_{e,\Delta} \setminus \bigcup_{i=1}^M R_i \right) \leq \eta.$$

The function Υ is differentiable on $(r, s) \in [0, 1]^2$. Moreover, it is not difficult to show that there exists $\xi < +\infty$ such that

$$\sup_{(r,s) \in [0,1]^2} \left| \det \left(\frac{\partial \Upsilon}{\partial (r,s)}(r, s) \right) \right| \leq \xi. \quad (\text{F.60})$$

Furthermore, for some ξ' , the map $\mathbf{w} \mapsto \mathbf{x}(\mathbf{w})$ satisfies

$$\left| \det \left(\frac{\partial}{\partial \mathbf{w}} (\varphi^{-1} \circ \mathbf{x})(\mathbf{w}) \right) \right| \leq \xi' \quad (\text{F.61})$$

for all \mathbf{w} . So, noting that $\Xi_{e,\Delta} = \mathbf{x}(\Upsilon[A_{e,\Delta}])$, we have

$$\mu_{\partial \mathcal{O}} \left(\Xi_{e,\Delta} \setminus \bigcup_{i=1}^M \mathbf{x}(\Upsilon[R_i]) \right) \leq \xi \xi' \eta. \quad (\text{F.62})$$

Choose $\eta = \frac{\varepsilon}{\xi \xi'}$.

Order all of the endpoints $s_{i,j}$ of the R_i , to produce $0 \leq s_1 < s_2 < \dots < s_N \leq 1$. Set

$$R'_{i,j} = R_j \cap ([s_i, s_{i+1}] \times [0, 1]). \quad (\text{F.63})$$

Set $\mathbf{a}_i = \mathbf{z}(s_i)$, $\mathbf{b}_i = \mathbf{z}(s_{i+1})$. Each $R'_{i,j}$ either has empty interior, or has the form $[s_i, s_{i+1}] \times [r_{i,j}, r_{i,j+1}]$. Hence, there exists a collection of disjoint intervals $[r_1, r_2], [r_3, r_4], \dots, [r_{2n_i-1}, r_{2n_i}]$ such that $\cup_j R'_{i,j} = \cup_j [s_i, s_{i+1}] \times [r_{2j-1}, r_{2j}]$. This collection of intervals has the desired properties. \square

¹⁴Because $\mathbf{n}^T \tilde{\mathbf{u}}(r) > 0$ for all $r \in [0, 1]$, the matrix $[\mathbf{V} \mid \tilde{\mathbf{u}}(r)]$ is full rank for all r , and hence for all r its smallest singular value is positive. Noting that $\sigma_{\min}(\mathbf{M})$ is a continuous function of \mathbf{M} and $\tilde{\mathbf{u}}(r)$ a continuous function of r , it must be that $\sigma_{\min}([\mathbf{V} \mid \tilde{\mathbf{u}}(r)])$ achieves its infimum over $[0, 1]$, and hence its infimum is strictly larger than zero.

G Proof of Lemma 5.4

Proof. Using the change of variables formula, it is not difficult to show that

$$\tilde{\nu}(\mathbf{x}) = 1 - \frac{1}{\pi} \int \frac{\langle \mathbf{n}(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \langle \mathbf{n}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle}{\|\mathbf{y} - \mathbf{x}\|_2^4} V(\mathbf{x}, \mathbf{y}) d\mu_{\partial\mathcal{O}}(\mathbf{y}). \quad (\text{G.1})$$

For any \mathbf{y} , we have

$$\begin{aligned} \int_{\mathbf{y}} \kappa(\mathbf{x}, \mathbf{y}) d\mu_{\partial\mathcal{O}}(\mathbf{y}) &= \frac{\rho(\mathbf{x})}{\pi} \int_{\mathbf{y}} \frac{\langle \mathbf{n}(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \langle \mathbf{n}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle}{\|\mathbf{x} - \mathbf{y}\|_2^4} V(\mathbf{x}, \mathbf{y}) d\mu_{\partial\mathcal{O}}(\mathbf{y}) \\ &= \rho(\mathbf{x})(1 - \tilde{\nu}(\mathbf{x})) \\ &\leq \rho_{\star}(1 - \nu_{\star}). \end{aligned} \quad (\text{G.2})$$

Similarly, for any $\mathbf{x} \in \partial\mathcal{O}$, we have

$$\begin{aligned} \int_{\mathbf{x}} \kappa(\mathbf{x}, \mathbf{y}) d\mu_{\partial\mathcal{O}}(\mathbf{x}) &= \frac{1}{\pi} \int_{\mathbf{x}} \rho(\mathbf{x}) \frac{\langle \mathbf{n}(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \langle \mathbf{n}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle}{\|\mathbf{x} - \mathbf{y}\|_2^4} V(\mathbf{x}, \mathbf{y}) d\mu_{\partial\mathcal{O}}(\mathbf{x}) \\ &\leq \frac{\rho_{\star}}{\pi} \int_{\mathbf{x}} \frac{\langle \mathbf{n}(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \langle \mathbf{n}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle}{\|\mathbf{x} - \mathbf{y}\|_2^4} V(\mathbf{x}, \mathbf{y}) d\mu_{\partial\mathcal{O}}(\mathbf{x}) \\ &\leq \rho_{\star}(1 - \tilde{\nu}(\mathbf{x})) \\ &\leq \rho_{\star}(1 - \nu_{\star}). \end{aligned} \quad (\text{G.3})$$

By Theorem II.1.6 of [Con90], this implies that for $g \in L^2[\partial\mathcal{O}]$, $\mathcal{T}[g] \in L^2[\partial\mathcal{O}]$, and $\|\mathcal{T}\|_{L^2 \rightarrow L^2} \leq \rho_{\star}(1 - \nu_{\star})$. \square

H Proof of Lemma 5.5

We can obtain Lemma 5.5 of the paper using the change of variables formula. Before jumping into the proof of this bound, we record a quick lemma:

Lemma H.1. *Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ($n > 1$) such that $\mathbf{u}^* \mathbf{v} \neq 0$. Then*

$$\left\| \mathbf{I} - \frac{\mathbf{u} \mathbf{v}^*}{\mathbf{u}^* \mathbf{v}} \right\| = \left\| \frac{\mathbf{u} \mathbf{v}^*}{\mathbf{u}^* \mathbf{v}} \right\| = \frac{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}{|\mathbf{u}^* \mathbf{v}|}. \quad (\text{H.1})$$

Proof. If \mathbf{u} and \mathbf{v} are linearly dependent, the result is immediate. Let us assume they are linearly independent. Let $\mathbf{M} = \mathbf{I} - \frac{\mathbf{u} \mathbf{v}^*}{\mathbf{u}^* \mathbf{v}}$, and consider the eigenvalues of

$$\mathbf{M} \mathbf{M}^* = \mathbf{I} - \frac{1}{\mathbf{u}^* \mathbf{v}} (\mathbf{u} \mathbf{v}^* + \mathbf{v} \mathbf{u}^*) + \frac{\mathbf{u} \mathbf{v}^* \mathbf{v} \mathbf{u}^*}{(\mathbf{u}^* \mathbf{v})^2} \quad (\text{H.2})$$

Notice that if $\mathbf{x} \perp \mathbf{u}, \mathbf{v}$, we have $\mathbf{M} \mathbf{M}^* \mathbf{x} = \mathbf{x}$. We can find an orthonormal basis of $n - 2$ vectors for $(\mathbf{u}, \mathbf{v})^\perp$. Since any orthonormal collection of eigenvectors of a symmetric matrix can be completed to an orthonormal basis of eigenvectors, there must exist two linearly independent eigenvectors lying in $\text{span}\{\mathbf{u}, \mathbf{v}\}$. If \mathbf{x} is an eigenvector with eigenvalue λ , we have

$$(1 - \lambda) \mathbf{x} = \left[\frac{1}{\mathbf{u}^* \mathbf{v}} (\mathbf{u} \mathbf{v}^* + \mathbf{v} \mathbf{u}^*) - \frac{\mathbf{u} \mathbf{v}^* \mathbf{v} \mathbf{u}^*}{(\mathbf{u}^* \mathbf{v})^2} \right] \mathbf{x}. \quad (\text{H.3})$$

For $\mathbf{x} \in \text{span}\{\mathbf{u}, \mathbf{v}\}$, write $\mathbf{x} = \alpha\mathbf{u} + \beta\mathbf{v}$. Plugging into the above equation and using linear independence of \mathbf{u} and \mathbf{v} , we obtain

$$(1 - \lambda)\alpha = \left(1 - \frac{\mathbf{v}^* \mathbf{v} \mathbf{u}^* \mathbf{u}}{(\mathbf{u}^* \mathbf{v})^2}\right) \alpha, \quad (\text{H.4})$$

$$(1 - \lambda)\beta = \frac{\mathbf{u}^* \mathbf{u}}{\mathbf{u}^* \mathbf{v}} \alpha + \beta. \quad (\text{H.5})$$

The first equation implies that either $\alpha = 0$, or $\lambda = \|\mathbf{v}\|_2^2 \|\mathbf{u}\|_2^2 / (\mathbf{u}^* \mathbf{v})^2$. If $\alpha = 0$, this implies that \mathbf{v} is an eigenvector with eigenvalue $\lambda = 1$. Plugging back into (H.3), and simplifying, we get $\mathbf{v} = \mathbf{0}$, contradicting $\mathbf{u}^* \mathbf{v} \neq 0$. Hence, $\alpha = 0$ cannot give an eigenvector under our assumptions, and it must be that the eigenvalue is $\lambda = \|\mathbf{v}\|_2^2 \|\mathbf{u}\|_2^2 / (\mathbf{u}^* \mathbf{v})^2$. By Cauchy-Schwarz, this quantity is strictly larger than one, and hence it is the largest eigenvalue of $\mathbf{M}\mathbf{M}^*$. Hence, $\|\mathbf{M}\| = \|\mathbf{v}\|_2 \|\mathbf{u}\|_2 / |\mathbf{u}^* \mathbf{v}|$. It is straightforward to observe that this quantity is also the norm of $\mathbf{u}\mathbf{v}^* / \mathbf{u}^* \mathbf{v}$. \square

Using this lemma and the change of variables formula, we can control the norm of the maps \mathcal{P}_i :

Lemma H.2. *For each i , $\|\mathcal{P}_i\|_{L^2 \rightarrow \mathbb{R}} \leq 2^{1/4} \beta f s / \ell$.*

Proof. We can define a restricted perspective projection $\tilde{\mathbf{p}} : \partial\mathcal{O}_+ \rightarrow \Pi_I$ via

$$\tilde{\mathbf{p}}(\mathbf{x}) = -f \frac{\mathbf{x}}{\langle \mathbf{x}, \mathbf{e}_3 \rangle}. \quad (\text{H.6})$$

Here, $\partial\mathcal{O}_+$ stands for the visible part of the object from the camera, defined as

$$\partial\mathcal{O}_+ = \{\mathbf{x} \in \partial\mathcal{O} \mid \text{conv}\{\mathbf{x}, \mathbf{0}\} \cap \mathcal{O} = \mathbf{x}\}.$$

The image coordinates are read off as the first two values $\mathbf{p}(\mathbf{x}) = \mathbf{P}_{12} \tilde{\mathbf{p}}(\mathbf{x})$, via

$$\mathbf{P}_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (\text{H.7})$$

The map \mathbf{p} is injective, and its inverse $\mathbf{p}^{-1} : \text{im}(\mathbf{p}) \subseteq \mathbb{R}^2 \rightarrow \partial\mathcal{O}_+$ exists. In our sensor model, we can write the value of the i -th pixel as

$$\begin{aligned} \mathcal{P}_i[g] &= \beta \int_{\mathbf{z} \in I_i \cap \text{im}(\mathbf{p})} g(\mathbf{p}^{-1} \mathbf{z}) \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|_2}, \mathbf{e}_3 \right\rangle^4 d\mu(\mathbf{z}) \\ &\doteq \beta \int_{\mathbf{z} \in I_i \cap \text{im}(\mathbf{p})} g(\mathbf{p}^{-1} \mathbf{z}) \cos^4(\alpha(\mathbf{z})) d\mu(\mathbf{z}). \end{aligned} \quad (\text{H.8})$$

We can change variables as above. Write

$$\mathcal{P}_i[g] = \beta \sum_j \int_{\mathbf{z} \in I_i \cap \mathbf{p}[\partial\mathcal{O}_+ \cap \varphi_j[U_j]]} g(\mathbf{p}^{-1} \mathbf{z}) \cos^4(\alpha(\mathbf{z})) d\mu(\mathbf{z}).$$

Here, $\varphi_j : U_j \rightarrow \Delta_j$ is defined as in Appendix C. Using the change of variables formula, this becomes

$$\mathcal{P}_i[g] = \beta \sum_j \int_{\mathbf{w} \in \varphi_j^{-1}[\partial\mathcal{O}_+ \cap \mathbf{p}^{-1}[I_i]]} g \circ \varphi_j(\mathbf{w}) \cos^4(\alpha(\mathbf{p} \varphi_j(\mathbf{w}))) \left| \det \left(\frac{\partial \mathbf{p} \circ \varphi_j}{\partial \mathbf{w}}(\mathbf{w}) \right) \right| d\mu(\mathbf{w}).$$

Writing

$$\zeta_j(\mathbf{z}) = \mathbb{1}_{\mathbf{z} \in I_i \cap \mathfrak{p}[\partial \mathcal{O}_+ \cap \varphi_j[U_j]]}. \quad (\text{H.9})$$

The above expression becomes

$$\mathcal{P}_i[g] = \beta \sum_j \int_{\mathbf{w} \in U_j} g \circ \varphi_j(\mathbf{w}) \zeta_j \circ \mathfrak{p} \circ \varphi_j(\mathbf{w}) \cos^4(\alpha(\mathfrak{p} \circ \varphi_j \mathbf{w})) \left| \det \left(\frac{\partial \mathfrak{p} \circ \varphi_j}{\partial \mathbf{w}}(\mathbf{w}) \right) \right| d\mu(\mathbf{w}). \quad (\text{H.10})$$

From this, we have

$$\|\mathcal{P}_i\|_{L^2 \rightarrow \mathbb{R}}^2 = \beta^2 \sum_j \int_{\mathbf{w} \in U_j} (\zeta_j \circ \mathfrak{p} \circ \varphi_j)^2(\mathbf{w}) \cos^8(\alpha(\mathfrak{p} \circ \varphi_j \mathbf{w})) \left| \det \left(\frac{\partial \mathfrak{p} \circ \varphi_j}{\partial \mathbf{w}}(\mathbf{w}) \right) \right|^2 d\mu(\mathbf{w}). \quad (\text{H.11})$$

To evaluate this, we can change variables again. Write

$$\begin{aligned} \|\mathcal{P}_i\|_{L^2 \rightarrow \mathbb{R}}^2 &= \beta^2 \sum_j \int_{\mathbf{z} \in \mathbb{R}^2} \zeta_j^2(\mathbf{z}) \cos^8(\alpha(\mathbf{z})) \left| \det \left(\frac{\partial \mathfrak{p} \circ \varphi_j}{\partial \mathbf{w}} \right) (\varphi_j^{-1} \mathfrak{p}^{-1} \mathbf{z}) \right|^2 \left| \det \left(\frac{\partial (\mathfrak{p} \circ \varphi_j)^{-1}}{\partial \mathbf{z}}(\mathbf{z}) \right) \right| d\mu(\mathbf{z}) \\ &= \beta^2 \sum_j \int_{\mathbf{z} \in \mathbb{R}^2} \zeta_j^2(\mathbf{z}) \cos^8(\alpha(\mathbf{z})) \left| \det \left(\frac{\partial \mathfrak{p} \circ \varphi_j}{\partial \mathbf{w}} \right) (\varphi_j^{-1} \mathfrak{p}^{-1} \mathbf{z}) \right| d\mu(\mathbf{z}). \end{aligned} \quad (\text{H.12})$$

To finish, we need to get a bound on the determinant term. For this, we use the fact that

$$\frac{\partial \tilde{\mathfrak{p}}}{\partial x} = -\frac{f}{\langle \mathbf{e}_3, \mathbf{x} \rangle} \left(\mathbf{I} - \frac{\mathbf{x} \mathbf{e}_3^*}{\langle \mathbf{e}_3, \mathbf{x} \rangle} \right), \quad (\text{H.13})$$

and

$$\begin{aligned} \det \left(\frac{\partial \mathfrak{p} \circ \varphi_j}{\partial \mathbf{w}} \right) &= \det \left(\mathbf{P}_{12} \frac{\partial \tilde{\mathfrak{p}}}{\partial \mathbf{x}} \mathbf{U}_j \right) \leq \left\| \mathbf{P}_{12} \frac{\partial \tilde{\mathfrak{p}}}{\partial \mathbf{x}} \mathbf{U}_j \right\|^2 \leq \left\| \frac{\partial \tilde{\mathfrak{p}}}{\partial \mathbf{x}} \right\|^2 \\ &= \frac{f^2}{(\mathbf{e}_3^* \mathbf{x})^2} \left\| \mathbf{I} - \frac{\mathbf{x} \mathbf{e}_3^*}{\mathbf{e}_3^* \mathbf{x}} \right\|^2 = \frac{f^2}{(\mathbf{e}_3^* \mathbf{x})^2} \left\| \frac{\mathbf{x}}{\mathbf{e}_3^* \mathbf{x}} \right\|_2^2, \end{aligned} \quad (\text{H.14})$$

where in the final line, we have used the above lemma. To bound the terms in this expression, notice that since $\mathbf{x} \in \mathcal{O}$, $\mathbf{e}_3^* \mathbf{x} \geq \ell$. Notice also that

$$-\frac{\mathbf{x}}{\mathbf{e}_3^* \mathbf{x}} = \frac{1}{f} \tilde{\mathfrak{p}}(\mathbf{x}). \quad (\text{H.15})$$

We have

$$\|\tilde{\mathfrak{p}}(\mathbf{x})\|_2 = \sqrt{f^2 + f^2 \tan^2 \alpha}. \quad (\text{H.16})$$

So, finally, we obtain

$$\left| \det \left(\frac{\partial \mathfrak{p} \circ \varphi_j}{\partial \mathbf{w}} \right) \right| \leq \frac{f^2}{\ell^2} (1 + \tan \alpha). \quad (\text{H.17})$$

It is not difficult to show¹⁵ that for all α ,

$$|\cos^8 \alpha \times (1 + \tan \alpha)| \leq \sqrt{2}. \quad (\text{H.18})$$

¹⁵In fact, the right hand side of (H.18) can be easily tightened to $\cos^8 \alpha \times (1 + \tan \alpha) \leq c < 1.1$. We will not pursue tight constants here, however.

Combining everything together, we obtain

$$\|\mathcal{P}_i\|_{L^2 \rightarrow \mathbb{R}}^2 \leq \sqrt{2} \frac{\beta^2 f^2}{\ell^2} \sum_j \int_{\mathbf{z}} \zeta_j^2(\mathbf{z}) d\mu(\mathbf{z}) \leq \sqrt{2} \frac{\beta^2 f^2}{\ell^2} s^2, \quad (\text{H.19})$$

giving the desired result. \square

This gives us a fairly direct proof of Lemma 5.5 of the paper:

Proof of Lemma 5.5. We have

$$\|\mathcal{P}\|_{L^2 \rightarrow \ell^2} = \sup_{\|g\|_{L^2} \leq 1} \left\| \begin{bmatrix} \mathcal{P}_1[g] \\ \vdots \\ \mathcal{P}_m[g] \end{bmatrix} \right\|_2. \quad (\text{H.20})$$

Because the I_i are disjoint and \mathbf{p} is injective, the sets $\Xi_i = \mathbf{p}^{-1}[I_i \cap \text{im}(\mathbf{p})]$ are disjoint, and

$$\|g\|_{L^2}^2 \geq \sum_i \|g \mathbb{1}_{\Xi_i}\|_{L^2}^2. \quad (\text{H.21})$$

From Lemma H.2, we have

$$|\mathcal{P}_i[g]| \leq \|\mathcal{P}_i[g]\|_{L^2 \rightarrow L^2} \|g \mathbb{1}_{\Xi_i}\|_{L^2} \leq (2^{1/4} \beta s f / \ell) \|g \mathbb{1}_{\Xi_i}\|_{L^2}, \quad (\text{H.22})$$

and so

$$\|\mathcal{P}[g]\|_{\ell^2}^2 \leq (2^{1/4} \beta s f / \ell)^2 \sum_i \|g \mathbb{1}_{\Xi_i}\|_{L^2}^2 \leq (2^{1/4} \beta s f / \ell)^2 \|g\|_{L^2}^2, \quad (\text{H.23})$$

completing the proof. \square

I Proofs from Section 6

Proof of Lemma 6.1.

Proof. Consider any $\widehat{\mathbf{A}} \in \Omega_1$. By the definition of Ω_1 , we have

$$\sup_{\mathbf{y} \in \text{cone}(\widehat{\mathbf{A}}), \|\mathbf{y}\| \leq 1} d(\mathbf{y}, \text{cone}(\widehat{\mathbf{A}})) \leq \gamma' \leq \frac{\gamma}{\gamma + 1} \leq \gamma.$$

We also have that

$$\sup_{\mathbf{y} \in \text{cone}(\widehat{\mathbf{A}}), \|\mathbf{y}\| \leq 1} d(\mathbf{y}, \text{cone}(\bar{\mathbf{A}})) \leq \max_{\mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq 1} \|\bar{\mathbf{A}}\mathbf{x} - \widehat{\mathbf{A}}\mathbf{x}\|_2.$$

Since $\forall \mathbf{x} \geq \mathbf{0}$, $\|\bar{\mathbf{A}}\mathbf{x} - \widehat{\mathbf{A}}\mathbf{x}\|_2 \leq \gamma' \|\bar{\mathbf{A}}\mathbf{x}\|_2$, we have

$$\begin{aligned} \left\{ \mathbf{x} \mid \mathbf{x} \geq \mathbf{0}, \|\widehat{\mathbf{A}}\mathbf{x}\|_2 \leq 1 \right\} &\subseteq \left\{ \mathbf{x} \mid \mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq 1 + \|\bar{\mathbf{A}}\mathbf{x} - \widehat{\mathbf{A}}\mathbf{x}\|_2 \right\} \\ &\subseteq \left\{ \mathbf{x} \mid \mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq \frac{1}{1 - \gamma'} \right\}. \end{aligned}$$

Thus

$$\begin{aligned}
\sup_{\mathbf{y} \in \text{cone}(\hat{\mathbf{A}}), \|\mathbf{y}\| \leq 1} d(\mathbf{y}, \text{cone}(\bar{\mathbf{A}})) &\leq \max_{\mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq 1} \|\bar{\mathbf{A}}\mathbf{x} - \hat{\mathbf{A}}\mathbf{x}\|_2 \\
&\leq \max_{\mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq \frac{1}{1-\gamma'}} \|\bar{\mathbf{A}}\mathbf{x} - \hat{\mathbf{A}}\mathbf{x}\|_2 \\
&= \frac{1}{1-\gamma'} \times \max_{\mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq 1} \|\bar{\mathbf{A}}\mathbf{x} - \hat{\mathbf{A}}\mathbf{x}\|_2 \\
&= \frac{\gamma'}{1-\gamma'} \leq \gamma.
\end{aligned}$$

Therefore, $\hat{\mathbf{A}} \in \Omega_1$ implies $\hat{\mathbf{A}} \in \Omega_0$ as desired. \square

Proof of Lemma 6.2.

Proof. By making the transformation $\mathbf{X} = \mathbf{x}\mathbf{x}^T$, we have that $\max_{\mathbf{x} \geq \mathbf{0}, \|\bar{\mathbf{A}}\mathbf{x}\|_2 \leq 1} \|\hat{\mathbf{A}}\mathbf{x} - \bar{\mathbf{A}}\mathbf{x}\|_2$ equals

$$\begin{aligned}
&\max \quad \left\langle \left(\hat{\mathbf{A}} - \bar{\mathbf{A}} \right)^T \left(\hat{\mathbf{A}} - \bar{\mathbf{A}} \right), \mathbf{X} \right\rangle \\
&\text{s.t.} \quad \left\langle \bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{X} \right\rangle \leq 1 \\
&\quad \mathbf{X} \geq \mathbf{0}, \mathbf{X} \succeq \mathbf{0} \\
&\quad \text{rank}(\mathbf{X}) = 1.
\end{aligned}$$

Dropping the rank constraint gives the result. \square

Proof of Theorem 6.3 We will prove the theorem via two lemmas below:

Lemma I.1. Consider $\Omega_3 \doteq \left\{ \hat{\mathbf{A}} \mid f(\hat{\mathbf{A}}) \leq \bar{\gamma} \right\}$, where $\bar{\gamma} = (\gamma')^2$ and

$$\begin{aligned}
f(\hat{\mathbf{A}}) &\doteq \min_{(\boldsymbol{\mu}, \beta)} \beta \\
&\text{s.t.} \quad \begin{bmatrix} \mathbf{I} & \hat{\mathbf{A}} - \bar{\mathbf{A}} \\ \left(\hat{\mathbf{A}} - \bar{\mathbf{A}} \right)^T & \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}, \beta \geq 0.
\end{aligned}$$

Then $\Omega_3 = \Omega_2$.

Proof. The above lemma follows directly from the fact that the dual problem of

$$\begin{aligned}
(\text{P}) \quad &\max \quad \left\langle \left(\hat{\mathbf{A}} - \bar{\mathbf{A}} \right)^T \left(\hat{\mathbf{A}} - \bar{\mathbf{A}} \right), \mathbf{X} \right\rangle \\
&\text{s.t.} \quad \left\langle \bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{X} \right\rangle \leq 1 \\
&\quad \mathbf{X} \geq \mathbf{0}, \mathbf{X} \succeq \mathbf{0},
\end{aligned} \tag{I.1}$$

can be written as

$$\begin{aligned}
\text{(D)} \quad & \min \quad \beta \\
& \text{s.t.} \quad \begin{bmatrix} \mathbf{I} & \hat{\mathbf{A}} - \bar{\mathbf{A}} \\ (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T & \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0} \\
& \quad \boldsymbol{\mu} \geq \mathbf{0}, \beta \geq 0,
\end{aligned}$$

with zero duality gap.

To derive that dual reformulation, we first reformulate problem (I.1) as:

$$\begin{aligned}
& \max \quad \left\langle (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}), \mathbf{X} \right\rangle \\
& \text{s.t.} \quad \left\langle \bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{X} \right\rangle - 1 \leq 0 \\
& \quad \mathbf{X} - \mathbf{Y} = \mathbf{0} \\
& \quad \mathbf{Y} \in \mathbb{R}_+^{n \times n}, \mathbf{X} \in S_+^n.
\end{aligned} \tag{I.2}$$

Let $\beta \in \mathbb{R}$ and $\boldsymbol{\mu} \in \mathbb{R}^{n \times n}$ correspond to the inequality constraint and equality constraint. Then the dual problem of (I.2) could be written in the following min-max form:

$$\begin{aligned}
& \min_{\substack{\beta \geq 0 \\ \boldsymbol{\mu} \text{ free}}} \max_{\substack{\mathbf{X} \in S_+^n \\ \mathbf{Y} \in \mathbb{R}_+^{n \times n}}} \left\langle (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}), \mathbf{X} \right\rangle - \beta \left(\left\langle \bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{X} \right\rangle - 1 \right) + \langle \boldsymbol{\mu}, \mathbf{X} - \mathbf{Y} \rangle. \tag{I.3}
\end{aligned}$$

By verifying that

$$\begin{aligned}
& \max_{\mathbf{X} \in S_+^n, \mathbf{Y} \in \mathbb{R}_+^{n \times n}} \left\langle (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}), \mathbf{X} \right\rangle - \beta \left(\left\langle \bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{X} \right\rangle - 1 \right) + \langle \boldsymbol{\mu}, \mathbf{X} - \mathbf{Y} \rangle \\
& = \max_{\mathbf{X} \in S_+^n, \mathbf{Y} \in \mathbb{R}_+^{n \times n}} \left\langle (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}) - \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} + \boldsymbol{\mu}, \mathbf{X} \right\rangle + \langle -\boldsymbol{\mu}, \mathbf{Y} \rangle + \beta \\
& = \begin{cases} \beta & \text{if } -(\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}) + \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \in S_+^n \text{ and } \boldsymbol{\mu} \in \mathbb{R}_+^{n \times n}, \\ +\infty & \text{otherwise} \end{cases},
\end{aligned}$$

we can write (I.3) as

$$\begin{aligned}
& \min \quad \beta \\
& \text{s.t.} \quad -(\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}) + \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \succeq \mathbf{0} \\
& \quad \boldsymbol{\mu} \geq \mathbf{0}, \beta \geq 0.
\end{aligned} \tag{I.4}$$

Because of Schur's complement, $-(\hat{\mathbf{A}} - \bar{\mathbf{A}})^T (\hat{\mathbf{A}} - \bar{\mathbf{A}}) + \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \succeq \mathbf{0}$ if and only if

$$\begin{bmatrix} \mathbf{I} & \hat{\mathbf{A}} - \bar{\mathbf{A}} \\ (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T & \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0}.$$

Thus (I.4) is equivalent to

$$\begin{aligned} \min \quad & \beta \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{I} & \hat{\mathbf{A}} - \bar{\mathbf{A}} \\ (\hat{\mathbf{A}} - \bar{\mathbf{A}})^T & \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0}. \\ & \boldsymbol{\mu} \geq \mathbf{0}, \beta \geq 0. \end{aligned}$$

Moreover, it can be easily verified that $\mathbf{X} = \frac{1}{2(\bar{\mathbf{A}}^T \bar{\mathbf{A}}, \mathbf{1}\mathbf{1}^T + \mathbf{I}_{n \times n})} (\mathbf{1}\mathbf{1}^T + \mathbf{I}_{n \times n})$ is a interior point in the feasible set of (I.1). Thus by Slater's condition, the duality gap is zero. \square

Hence, instead of solving (6.2), we can work with

$$\min \left\{ \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \mid \mathbf{L} + \mathbf{S} = \hat{\mathbf{A}}, \hat{\mathbf{A}} \in \Omega_3 \right\}.$$

The following lemma completes our proof of Theorem 6.3:

Lemma I.2. *Our relaxed convex optimization problem*

$$\min \left\{ \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \mid \mathbf{L} + \mathbf{S} = \hat{\mathbf{A}}, \hat{\mathbf{A}} \in \Omega_3 \right\} \quad (\text{I.5})$$

is equivalent to problem (6.4).

Proof. Problem (I.5)

$$\begin{aligned} \min \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \mathbf{L} + \mathbf{S} = \hat{\mathbf{A}} \\ & f(\hat{\mathbf{A}}) \leq \bar{\gamma}, \end{aligned}$$

can be easily written as

$$\begin{aligned} \min \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{I} & \mathbf{L} + \mathbf{S} - \bar{\mathbf{A}} \\ (\mathbf{L} + \mathbf{S} - \bar{\mathbf{A}})^T & \beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0} \\ & \boldsymbol{\mu} \geq \mathbf{0}, \bar{\gamma} \geq \beta \geq 0. \end{aligned} \quad (\text{I.6})$$

Whenever $(\mathbf{L}^*, \mathbf{S}^*, \beta^*, \boldsymbol{\mu}^*)$ is an optimal solution to problem (I.6), $(\mathbf{L}^*, \mathbf{S}^*, \bar{\gamma}, \boldsymbol{\mu}^*)$ is still feasible by noting that

$$\begin{bmatrix} \mathbf{I} & \mathbf{L}^* + \mathbf{S}^* - \bar{\mathbf{A}} \\ (\mathbf{L}^* + \mathbf{S}^* - \bar{\mathbf{A}})^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{L}^* + \mathbf{S}^* - \bar{\mathbf{A}} \\ (\mathbf{L}^* + \mathbf{S}^* - \bar{\mathbf{A}})^T & \beta^* \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu}^* \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\bar{\gamma} - \beta^*) \bar{\mathbf{A}}^T \bar{\mathbf{A}} \end{bmatrix} \succeq \mathbf{0}.$$

Moreover, the objective value does not change. Thus $(\mathbf{L}^*, \mathbf{S}^*, \bar{\gamma}, \boldsymbol{\mu}^*)$ is also an optimal solution. Therefore, we can rewrite problem (I.6) as

$$\begin{aligned} \min \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{I} & \mathbf{L} + \mathbf{S} - \bar{\mathbf{A}} \\ (\mathbf{L} + \mathbf{S} - \bar{\mathbf{A}})^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu} \end{bmatrix} \succeq \mathbf{0} \\ & \boldsymbol{\mu} \geq \mathbf{0}, \end{aligned}$$

□

J Scalable Complexity Reduction using L-ADMM

Per Nesterov's advice that "...the proper use of the problem's structure can lead to efficient optimization methods..." [Nes07], we would like to search for a scalable algorithm that takes full advantage of the structure of (6.4). The problem can be rephrased as

$$\begin{aligned} \min \quad & \|L\|_* + \lambda \|S\|_1 + \mathcal{I}(Z \succeq 0) + \mathcal{I}(\mu \geq 0) \\ \text{s.t.} \quad & Z - \begin{bmatrix} I & L + S - \bar{A} \\ (L + S - \bar{A})^T & \bar{\gamma} \bar{A}^T \bar{A} - \mu \end{bmatrix} = 0, \end{aligned} \quad (\text{J.1})$$

where the indicator function $\mathcal{I}(x \in \mathcal{X})$ is defined as

$$\mathcal{I}(x \in \mathcal{X}) = \begin{cases} 0, & \text{if } x \in \mathcal{X} \\ +\infty, & \text{otherwise.} \end{cases}$$

The most important structure in (J.1) seems to be that the objective function and constraints are *separable*. This naturally suggests the use of alternating directions methods. We will adopt the recently proposed Linearized Alternating Direction Method of Multipliers (L-ADMM) [ZBBO10, ZBO11, MXZ12] for (J.1). The L-ADMM is well adapted for problems of this form (J.1). This method works with the Augmented Lagrangian,

$$\begin{aligned} \mathcal{L}_\rho(Z, L, S, \mu; Y) \quad & \doteq \quad \|L\|_* + \lambda \|S\|_1 + \mathcal{I}(Z \succeq 0) + \mathcal{I}(\mu \geq 0) \\ & + \left\langle Y, Z - \begin{bmatrix} I & L + S - \bar{A} \\ (L + S - \bar{A})^T & \bar{\gamma} \bar{A}^T \bar{A} - \mu \end{bmatrix} \right\rangle \\ & + \frac{\rho}{2} \left\| Z - \begin{bmatrix} I & L + S - \bar{A} \\ (L + S - \bar{A})^T & \bar{\gamma} \bar{A}^T \bar{A} - \mu \end{bmatrix} \right\|_F^2. \end{aligned}$$

Here, Y is the multiplier of the linear constraint, and $\rho > 0$ is the penalty parameter. For notational convenience, partition Z as

$$Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix},$$

in accordance with the block structure of

$$\begin{bmatrix} I & L + S - \bar{A} \\ (L + S - \bar{A})^T & \bar{\gamma} \bar{A}^T \bar{A} - \mu \end{bmatrix}.$$

Following the same rule, we define Y_{11} , Y_{12} , Y_{21} and Y_{22} correspondingly.

The L-ADMM algorithm, operating on $\mathcal{L}_\rho(Z, L, S, \mu; Y)$, consists of the following three steps:

1. Minimize $\mathcal{L}_\rho(\mathbf{Z}, \mathbf{L}, \mathbf{S}, \boldsymbol{\mu}; \mathbf{Y})$ with respect to \mathbf{Z} , while keeping all the other variables fixed:

$$\begin{aligned}
\mathbf{Z}^{k+1} &= \arg \min_{\mathbf{Z}} \mathcal{L}_\rho(\mathbf{Z}, \mathbf{L}^k, \mathbf{S}^k, \boldsymbol{\mu}^k; \mathbf{Y}^k) \\
&= \arg \min_{\mathbf{Z}} \mathcal{I}(\mathbf{Z} \succeq \mathbf{0}) + \frac{\rho}{2} \left\| \mathbf{Z} - \left(\begin{bmatrix} \mathbf{I} & \mathbf{L}^k + \mathbf{S}^k - \bar{\mathbf{A}} \\ (\mathbf{L}^k + \mathbf{S}^k - \bar{\mathbf{A}})^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu}^k \end{bmatrix} - \frac{\mathbf{Y}^k}{\rho} \right) \right\|_F^2 \\
&= \mathbf{Q}^k (\boldsymbol{\Lambda}^k)_+ (\mathbf{Q}^k)^T,
\end{aligned}$$

where $\mathbf{Q}^k \boldsymbol{\Lambda}^k (\mathbf{Q}^k)^T$ is any eigenvalue decomposition of

$$\begin{bmatrix} \mathbf{I} & \mathbf{L}^k + \mathbf{S}^k - \bar{\mathbf{A}} \\ (\mathbf{L}^k + \mathbf{S}^k - \bar{\mathbf{A}})^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu}^k \end{bmatrix} - \frac{\mathbf{Y}^k}{\rho},$$

and $\boldsymbol{\Lambda}^k = \text{diag}(\{\lambda_i^k\}_{i=1}^{m+n})$ and $(\boldsymbol{\Lambda}^k)_+ = \text{diag}(\{\max(\lambda_i^k, 0)\}_{i=1}^{m+n})$.

2. Fix variables \mathbf{Z} and \mathbf{Y} and update \mathbf{L} , \mathbf{S} and $\boldsymbol{\mu}$. Instead of minimizing $\mathcal{L}_\rho(\mathbf{Z}^{k+1}, \mathbf{L}, \mathbf{S}, \boldsymbol{\mu}; \mathbf{Y}^k)$ directly, we construct a surrogate function $\widehat{\mathcal{L}}_\rho(\mathbf{Z}^{k+1}, \mathbf{L}, \mathbf{S}, \boldsymbol{\mu}; \mathbf{Y}^k)$ by linearizing $\mathcal{L}_\rho(\mathbf{Z}^{k+1}, \mathbf{L}, \mathbf{S}, \boldsymbol{\mu}; \mathbf{Y}^k)$, and then minimize $\widehat{\mathcal{L}}_\rho(\mathbf{Z}^{k+1}, \mathbf{L}, \mathbf{S}, \boldsymbol{\mu}; \mathbf{Y}^k)$ in $(\mathbf{L}, \mathbf{S}, \boldsymbol{\mu})$ -direction. This gives

$$\begin{pmatrix} \mathbf{L}^{k+1} \\ \mathbf{S}^{k+1} \\ \boldsymbol{\mu}^{k+1} \end{pmatrix} = \arg \min_{(\mathbf{L}, \mathbf{S}, \boldsymbol{\mu})} \widehat{\mathcal{L}}_\rho(\mathbf{Z}^{k+1}, \mathbf{L}, \mathbf{S}, \boldsymbol{\mu}; \mathbf{Y}^k) = \begin{pmatrix} \arg \min_{\mathbf{L}} \left(\|\mathbf{L}\|_* + \frac{\rho}{2\tau} \|\mathbf{L} - \mathbf{F}^k\|_F^2 \right) \\ \arg \min_{\mathbf{S}} \left(\lambda \|\mathbf{S}\|_1 + \frac{\rho}{2\tau} \|\mathbf{S} - \mathbf{G}^k\|_F^2 \right) \\ \arg \min_{\boldsymbol{\mu}} \left(\mathcal{I}(\boldsymbol{\mu} \geq \mathbf{0}) + \frac{\rho}{2\tau} \|\boldsymbol{\mu} - \mathbf{K}^k\|_F^2 \right) \end{pmatrix}$$

where

$$\begin{aligned}
\mathbf{F}^k &:= \mathbf{L}^k + 2\tau \left(\mathbf{Z}_{12}^{k+1} - \mathbf{L}^k - \mathbf{S}^k + \bar{\mathbf{A}} + \mathbf{Y}_{12}^k / \rho \right) \\
\mathbf{G}^k &:= \mathbf{S}^k + 2\tau \left(\mathbf{Z}_{12}^{k+1} - \mathbf{L}^k - \mathbf{S}^k + \bar{\mathbf{A}} + \mathbf{Y}_{12}^k / \rho \right) \\
\mathbf{K}^k &:= \boldsymbol{\mu}^k - \tau \left(\mathbf{Z}_{22}^{k+1} - \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} + \boldsymbol{\mu} + \mathbf{Y}_{22}^k / \rho \right)
\end{aligned}$$

and $\tau \geq 0$ is a given step size to be discussed later. The advantage of this linearization is that the sub-minimizations over \mathbf{L} , \mathbf{S} , and $\boldsymbol{\mu}$ have efficient, closed-form solutions. Let $\mathcal{S}_\theta : \mathbb{R} \rightarrow \mathbb{R}$ denote the shrinkage operator

$$\mathcal{S}_\theta(x) = \text{sgn}(x) \max(|x| - \theta, 0)$$

and extend it to matrices by applying it componentwise. It is easy to show that $\mathbf{S}^{k+1} = \mathcal{S}_{\tau\lambda/\rho}(\mathbf{G}^k)$. Similarly, for any matrix \mathbf{X} , denote \mathcal{D}_θ as the singular value thresholding operator $\mathcal{D}_\theta(\mathbf{X}) = \mathbf{U} \mathcal{S}_\theta(\boldsymbol{\Sigma}) \mathbf{V}^*$ where $\mathbf{X} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^*$ is any singular value decomposition. It is not difficult to show that $\mathbf{L}^{k+1} = \mathcal{D}_{\tau/\rho}(\mathbf{F}^k)$. For $\boldsymbol{\mu}^{k+1}$, we simply have $\boldsymbol{\mu}^{k+1} = (\mathbf{K}^k)_+ = [\max(\mathbf{K}_{ij}^k, 0)]_{ij}$.

3. Update the dual multiplier

$$\mathbf{Y}^{k+1} = \mathbf{Y}^k + \rho \left(\mathbf{Z}^{k+1} - \begin{bmatrix} \mathbf{I} & \mathbf{L}^{k+1} + \mathbf{S}^{k+1} - \bar{\mathbf{A}} \\ \left(\mathbf{L}^{k+1} + \mathbf{S}^{k+1} - \bar{\mathbf{A}} \right)^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu}^{k+1} \end{bmatrix} \right)$$

Putting these results together, we obtain an L-ADMM iterates as follows:

Algorithm 1 L-ADMM for (J.1)

Step 1. Generate \mathbf{Z}^{k+1} :

$$\mathbf{Z}^{k+1} = \mathbf{Q}^k \left(\boldsymbol{\Lambda}^k \right)_+ \left(\mathbf{Q}^k \right)^T.$$

Step 2. Generate \mathbf{L}^{k+1} , \mathbf{S}^{k+1} and $\boldsymbol{\mu}^{k+1}$:

$$\begin{cases} \mathbf{L}^{k+1} = \mathcal{D}_{\tau/\rho} \left(\mathbf{F}^k \right) \\ \mathbf{S}^{k+1} = \mathcal{S}_{\tau\lambda/\rho} \left(\mathbf{G}^k \right) \\ \boldsymbol{\mu}^{k+1} = \left(\mathbf{K}^k \right)_+ \end{cases}$$

Step 3. Update the multiplier \mathbf{Y}^{k+1} :

$$\mathbf{Y}^{k+1} = \mathbf{Y}^k + \rho \left(\mathbf{Z}^{k+1} - \begin{bmatrix} \mathbf{I} & \mathbf{L}^{k+1} + \mathbf{S}^{k+1} - \bar{\mathbf{A}} \\ \left(\mathbf{L}^{k+1} + \mathbf{S}^{k+1} - \bar{\mathbf{A}} \right)^T & \bar{\gamma} \bar{\mathbf{A}}^T \bar{\mathbf{A}} - \boldsymbol{\mu}^{k+1} \end{bmatrix} \right).$$

It can be shown that, with a proper choice of τ , our L-ADMM algorithm converges globally with rate $O(1/k)$. The proper τ is dictated by the following lemma, which bounds the norm of the operator in the linear constraint in (J.1):

Lemma J.1. Let $\mathcal{G} : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{(m+n) \times (m+n)}$ such that

$$\mathcal{G}(\mathbf{L}, \mathbf{S}, \boldsymbol{\mu}) := \begin{bmatrix} \mathbf{0} & \mathbf{L} + \mathbf{S} \\ (\mathbf{L} + \mathbf{S})^T & -\boldsymbol{\mu} \end{bmatrix}.$$

Then we have operator norm $\|\mathcal{G}\| = 2$.

Proof. Set $\mathcal{G}_1(\mathbf{L} + \mathbf{S}) \doteq \begin{bmatrix} \mathbf{0} & \mathbf{L} + \mathbf{S} \\ (\mathbf{L} + \mathbf{S})^T & \mathbf{0} \end{bmatrix}$ and $\mathcal{G}_2(\boldsymbol{\mu}) \doteq \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\mu} \end{bmatrix}$. It can be easily verified

that operator norms $\|\mathcal{G}_1\| = 2$ and $\|\mathcal{G}_2\| = 1$. Then we have

$$\begin{aligned}
\|\mathcal{G}\|^2 &= \sup_{\|\boldsymbol{\mu}\|_F^2 + \|(\mathbf{L}, \mathbf{S})\|_F^2 = 1} \left\| \begin{bmatrix} \mathbf{0} & \mathbf{L} + \mathbf{S} \\ (\mathbf{L} + \mathbf{S})^T & -\boldsymbol{\mu} \end{bmatrix} \right\|_F^2 \\
&= \sup_{\|\boldsymbol{\mu}\|_F^2 + \|(\mathbf{L}, \mathbf{S})\|_F^2 = 1} \left(\left\| \begin{bmatrix} \mathbf{0} & \mathbf{L} + \mathbf{S} \\ (\mathbf{L} + \mathbf{S})^T & \mathbf{0} \end{bmatrix} \right\|_F^2 + \left\| \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\mu} \end{bmatrix} \right\|_F^2 \right) \\
&= \sup_{\|\boldsymbol{\mu}\|_F^2 + \|(\mathbf{L}, \mathbf{S})\|_F^2 = 1} (\|\mathcal{G}_1\|^2 \|(\mathbf{L}, \mathbf{S})\|_F^2 + \|\mathcal{G}_2\|^2 \|\boldsymbol{\mu}\|_F^2) \\
&= \max(\|\mathcal{G}_1\|^2, \|\mathcal{G}_2\|^2) \\
&= 4,
\end{aligned}$$

□

Combining it with convergence results from Appendix A of [MXZ12], and results on convergence rate from [HY12] (Theorem 4.1), we obtain the following convergence guarantee for our algorithm:

Theorem J.2. (*Convergence Results*) Suppose $0 < \tau < 0.25$. Then the sequence $\left\{ \left(\mathbf{Z}^k, \mathbf{L}^k, \mathbf{S}^k, \boldsymbol{\mu}^k \right) \right\}$ produced by Alg-1 from any starting point converges to an optimal solution with rate $O(1/k)$.