The Role of Functional Analysis in Global Illumination

James Arvo

Program of Computer Graphics
Cornell University
Ithaca, NY 14853

Abstract: The problem of global illumination is virtually synonymous with solving the rendering equation. Although a great deal of research has been directed toward Monte Carlo and finite element methods for solving the rendering equation, little is known about the continuous equation beyond the existence and uniqueness of its solution. The continuous problem may be posed in terms of linear operators acting on infinite-dimensional function spaces. Such operators are fundamentally different from their finite-dimensional counterparts, and are properly studied using the methods of functional analysis. This paper summarizes some of the basic concepts of functional analysis and shows how these concepts may be applied to a linear operator formulation of the rendering equation. In particular, operator norms are obtained from thermodynamic principles, and a number of common function spaces are shown to be closed under global illumination. Finally, several fundamental operators that arise in global illumination are shown to be nearly finite-dimensional in that they can be uniformly approximated by matrices.

1 Introduction

The rendering equation has served to codify the study of global illumination since its introduction by Kajiya in 1986 [12]. Methods for the construction and solution of discrete versions of the rendering equation have been studied extensively; however, the continuous equation and its solution have received relatively little attention. This emphasis is somewhat justified given that digital computation rests upon finite representations; even though a solution exists within an infinite-dimensional function space, its approximation must be representable by a finite machine in finite time. Nevertheless, the fidelity of the finite-dimensional approximation can only be fully understood in the context of the infinite-dimensional problem.

A natural way to approach the infinite-dimensional problem is through the formalism of functional analysis. Functional analysis is the study of algebraic and topological properties of abstract spaces, particularly infinite-dimensional spaces of functions and operators that transform them. This level of abstraction can provide valuable insight by characterizing both the operators and the solutions of operator equations, and is particularly well-suited to the study of integral equations.

The techniques of functional analysis draw upon measure theory and topology. Measure theory provides the basis for integration while topology formalizes the notions of "closeness" and "convergence" that underly all of analysis. These concepts are extremely general and extend to infinite-dimensional spaces, where familiar notions of Euclidean

space often break down. For example, a linear operator on an infinite-dimensional space may be discontinuous, which is impossible in $\mathbb{R}^n$. Also, different norms on these spaces need not be equivalent; that is, a sequence may converge with respect to one norm, but not another. In more abstract terms, this means that different norms can induce different topologies. As a final example, the spectrum of an infinite-dimensional linear operator is generally more complex than that of a matrix.

The remainder of the paper is a brief exploration of these concepts and a set of guideposts for applying them in the context of global illumination. Pointers into the relevant literature are provided for those who wish to learn more.

2 The Classical Transport Equation

Equations governing the transport of radiant energy have appeared in astrophysics [5], thermal engineering [22], and illumination engineering [16]. For global illumination the governing equation may be expressed in terms of an unknown *surface radiance* function f resulting from a source term f_0. In continuous form, the governing equation is

$$f(\mathbf{r},\mathbf{u}) = f_0(\mathbf{r},\mathbf{u}) + \int_{\Omega_i} k(\mathbf{r};\mathbf{u}' \to \mathbf{u})\, f(\mathbf{r}',\mathbf{u}')\, d\mu(\mathbf{u}'), \tag{1}$$

where Ω_i denotes the hemisphere of incoming directions with respect to the surface point $\mathbf{r}$, and $d\mu$ denotes cosine-weighted integration over directions. The kernel k is the bidirectional reflectance distribution function expressed in terms of global coordinates, and $\mathbf{r}'$ is the point that is visible to $\mathbf{r}$ in the direction $-\mathbf{u}'$. Equation (1) is essentially the formulation posed by Polyak [17] in thermal engineering, which can be used to derive Kajiya's rendering equation by a change of variables.

Equation (1) is often referred to as a Fredholm integral equation of the second kind although it has several features that do not conform to the standard definition. First, the integral for a given point $\mathbf{r}$ does not depend on the entire domain of f, but only a vanishingly small subset; a difference that disappears when one considers only radiosity. Secondly, $\mathbf{r}'$ is an implicit function that depends on the argument $\mathbf{r}$ and the dummy variable $\mathbf{u}'$. This seemingly minor difference is responsible for the most important distinguishing feature of radiative transfer problems; the non-localness of the interactions. However, this property is difficult to work with as it appears above.

To precisely define the implicit function $\mathbf{r}'$ we introduce two intermediate functions. First, let $\mathcal{M}$ denote the collection of all surfaces in an environment, and define the *visible-surface function* $\nu(\mathbf{r},\mathbf{u})$ by

$$\nu(\mathbf{r},\mathbf{u}) \;\equiv\; \inf\{x > 0 : \mathbf{r} + x\mathbf{u} \in \mathcal{M}\},$$

where "$\inf$" means greatest lower bound. This function returns the distance from $\mathbf{r}$ to the nearest point on $\mathcal{M}$ in the direction of $\mathbf{u}$ (Glassner [9] gives a similar definition). If no such point exists, we define $\nu(\mathbf{r},\mathbf{u}) = \infty$. We then define the *ray casting function* by

$$\mathbf{p}(\mathbf{r},\mathbf{u}) \;\equiv\; \mathbf{r} + \nu(\mathbf{r},\mathbf{u})\mathbf{u},$$

which is the point of intersection with $\mathcal{M}$ when one exists; the function is undefined otherwise. It follows that $\mathbf{r}'(\mathbf{r},\mathbf{u}) \;\equiv\; \mathbf{p}(\mathbf{r},-\mathbf{u})$. This function will also be useful in recasting equation (1) into operator form.

3 Function Spaces

The formal study of equation (1) must include a characterization of the class of admissible functions for f_0. Let X denote a vector space of real-valued surface radiance functions defined on $\mathcal{M} \times \mathcal{S}^2$, where $\mathcal{S}^2$ is the unit sphere. For purposes of analysis, the space X must have additional structure; minimally, it must have a *topology*, which defines open sets or *neighborhoods* of functions [20]. Here we shall only consider the *metric topologies* that are an automatic consequence of a norm defined on the space. Infinitely many norms can be defined on the space of radiance functions, each inducing a different topology and each providing a different notion of size, distance, and convergence.

A second useful feature for the space X is that of *completeness*. For a normed space, completeness means that all *Cauchy sequences* converge. That is, if $f_1, f_2, \ldots$ is a sequence of functions in X such that

$$\lim_{n,m \to \infty} || f_n - f_m || = 0,$$

then there exists a function $f \in X$ that the sequence converges to. Since the Cauchy sequence makes no reference to the limit point itself, this definition is an intrinsic property of the space X; it ensures that every function that can be approximated by elements of the space is itself in the space. A vector space that is complete with respect to a given norm is called a *Banach space* [25, 15].

3.1 Measures and Lebesgue Integration

A *measure* is mechanism for generalizing the notion of area or volume to abstract sets. A measure is a countably additive positive set function; that is, if m is a measure over a set S, then $0 \leq m(A_1 \cup A_2 \cup \cdots) = m(A_1) + m(A_2) + \cdots$ whenever the sets $A_1, A_2, \ldots$ are mutually disjoint subsets of S. We shall let σ denote the canonical measure on the sphere [3, p. 276]; that is, $\sigma(A)$ is the surface area of any (measurable) subset $A \subset \mathcal{S}^2$.

The Lebesgue integral is based upon the concept of measure and is commonly denoted as in equation (1), where $d\mu(\mathbf{u}')$[1] signifies both the measure μ and the dummy variable $\mathbf{u}'$. The Lebesgue integral is an appropriate abstraction for functional analysis, where the emphasis is on integrals as transformations rather than the numerical aspects of integration. While the simpler concept of the Riemann integral is equivalent when it exists, the Lebesgue integral is more robust with respect to limiting operations, which makes it a valuable tool for defining abstract spaces such as Banach spaces [21].

The measure-theoretic notation used in (1) also allows us to introduce the convenient measure μ that incorporates a cosine weighting. Formally, the measure μ is defined by

$$\mu(E) \equiv \int_E |\mathbf{u} \cdot \mathbf{n}(\mathbf{r})| \; d\sigma(\mathbf{u}), \tag{2}$$

where $E \subset \mathcal{S}^2$ and $\mathbf{n}(\mathbf{r})$ is the surface normal at the point $\mathbf{r} \in \mathcal{M}$; thus, μ implicitly depends on $\mathbf{r}$. Equation (2) shows how one measure may be defined in terms of another. The new measure μ eliminates the proliferation of cosine factors that appear in radiative

[1] This notation differs among authors. Alternatives include $d\mu$ [20, p. 20] and $\mu(d\mathbf{u}')$ [7, p. 95].

transfer computations, but more importantly, it emphasizes that the cosine is an artifact of surface integration. This observation is vital in defining appropriate norms and inner products for radiance functions.

3.2 The L_p Spaces

The Lebesgue integral provides a mechanism for defining a family of fundamental function spaces called *Lebesgue spaces* or L_p-spaces [20, p. 66]. These spaces are defined in terms of the L_p-norms, which for radiance functions are given by

$$|| f ||_p \equiv \left[\int_{\mathcal{M}} \int_{\mathcal{S}^2} | f(\mathbf{r}, \mathbf{u}) |^p \, d\mu(\mathbf{u}) \, dm(\mathbf{r}) \right]^{1/p}, \tag{3}$$

where m denotes area measure over $\mathcal{M}$, and $1 \leq p \leq \infty$ is a real number[2]. The collection of functions with finite L_p-norm is denoted by $L_p(\mu \times m)$, or simply by L_p if the measure is clear from context. Each L_p-space is a Banach space because the limit of L_p functions is again in L_p. The L_1-norm of a radiance function f is its power, and therefore carries the corresponding units [watts].

As p approaches infinity, the L_p-norm places greater emphasis on the largest values attained by the function. In the limit, when $p = \infty$, the L_p-norm becomes

$$\| f \|_\infty \equiv \operatorname*{ess\,sup}_{\mathbf{r} \in \mathcal{M}} \operatorname*{ess\,sup}_{\mathbf{u} \in \mathcal{S}^2} | f(\mathbf{r}, \mathbf{u}) | , \tag{4}$$

where "ess sup" is the *essential supremum*, the least upper bound obtainable by ignoring a subset of measure zero. More formally, if h is a real function defined on a set A, then

$$\operatorname*{ess\,sup}_{x \in A} h(x) \;\equiv\; \inf \{ \, m \mid h(x) \leq m \;\text{ for almost every }\; x \in A \, \} , \tag{5}$$

where "almost" means to within a set of measure zero [25]. The essential supremum corresponds to the notion of maximum, but with two important differences. First, it is the *least upper bound*, or *supremum*, which exists for all bounded subsets of $\mathbb{R}$. Secondly, it is unaffected by values that cannot contribute to the integral of the function; this includes values assumed at isolated points or over sets of measure zero, such as curves in $\mathbb{R}^2$.

The L_∞-norm also has a direct interpretation; it is the maximum radiance attained or approached over all surfaces and directions; it follows that its units are those of radiance [watts/m^{2}sr]. Finally, the L_2-norm is important since it has additional algebraic properties that make it a Hilbert space, as described in the next section.

3.3 The Hilbert Space L_2

A *Hilbert space* $\mathcal{H}$ is a Banach space whose norm is defined in terms of an inner product $\langle \cdot \mid \cdot \rangle$ defined on $\mathcal{H} \times \mathcal{H}$ [15]. More precisely, the norm $|| \cdot ||$ is defined by

$$|| x || \equiv \sqrt{\langle x \mid x \rangle}, \tag{6}$$

[2] The definition is also meaningful for $p < 1$, but these norms are not *strictly convex* and are often excluded [14].

where $x \in \mathcal{H}$. Not every Banach space is a Hilbert space. Equation (6) imbues the norm and the entire space with many useful properties that a general Banach space does not enjoy: for example, the concept of orthogonality is given meaning, and every continuous linear functional (scalar-valued function) h defined on $\mathcal{H}$ can be represented by an inner product; that is, $h(x) = \langle\, x \mid y \,\rangle$ for some $y \in \mathcal{H}$.

While L_1 and L_∞ are useful Banach spaces, their norms are not compatible with any inner product, and therefore they are not Hilbert spaces. This fact is easily demonstrated using a clever observation attributed to John von Neumann. The observation is that the *parallelogram law*

$$\| f - g \|^2 + \| f + g \|^2 \;=\; 2 \| f \|^2 \;+\; 2 \| g \|^2 \tag{7}$$

is satisfied for all f and g in a Banach space if and only if it is also a Hilbert space [18, p. 27]. This is easy to verify using equation (6) and the bilinearity of inner products. However, equation (7) is violated by both the L_1-norm and the L_∞-norm; for instance, simply let f and g be disjoint unit step-functions.

The space L_2, on the other hand, is a Hilbert space with the inner product

$$\langle\, f \mid g \,\rangle \equiv \int_{\mathcal{M}} \int_{\mathcal{S}^2} f(\mathbf{r}, \mathbf{u})\, g(\mathbf{r}, \mathbf{u})\; d\mu(\mathbf{u})\; dm(\mathbf{r}). \tag{8}$$

In fact, the space L_2 is the only L_p-space that is also a Hilbert space. This can be shown by observing that L_p is isomorphic to its own *dual space* [15, p. 106] only when $p = 2$; this is a feature of all Hilbert spaces.

The distinction between a Banach space and a Hilbert space can be important in global illumination; for example, Galerkin-based methods [26] depend on orthogonal projections, which are only meaningful in a Hilbert space.

4 Linear Operators

A function whose domain is a function space such as L_p is commonly referred to as an *operator*. Since all of the L_p-spaces are infinite-dimensional for typical measures, we often speak of the operators defined on them as being infinite-dimensional as well. Common infinite-dimensional linear operators include *integral* or *differential* operators, *projection* operators, and the *multiplication* operator defined by

$$(\mathbf{M}_g\, f)(x) \;\equiv\; f(x)\, g(x), \tag{9}$$

where g is some function. An operator is finite-dimensional, or *degenerate*, only if its range is finite-dimensional. An important type of degenerate operator is one that projects onto a subspace spanned by a finite number of basis functions.

The collection of linear operators that map one function space into another is itself a vector space under pointwise addition and scalar multiplication. Such a vector space can be made into a Banach space by imposing a norm. For instance, the standard *operator norm* corresponding to a given function norm $||\cdot||$ is defined by

$$||\, \mathbf{A} \,|| \equiv \inf \{\, m \;:\; ||\, \mathbf{A}h \,|| \;\leq\; m\, ||\, h \,||\;\text{ for all }\; h \,\}\,. \tag{10}$$

Operator equations are a generalization of integral equations in which the essential algebraic properties such as linearity and associativity are preserved, while less important details of the operators are hidden. Operator equations were first applied in global illumination by Kajiya [12], where they were used to show the existence and uniqueness of solutions to the rendering equation.

4.1 Features of Infinite-Dimensional Operators

Linear operators can possess properties that do not occur in their finite-dimensional counterparts, which correspond to matrices. For example, a linear operator may be *unbounded*; that is, it may have infinite norm. The simplest example of such an operator is the point-evaluation functional $\phi_x(f) = f(x)$, which is unbounded on all L_p-spaces. Also, most differential operators are unbounded on spaces of differentiable functions. For linear operators, unboundedness is synonymous with being discontinuous.

Another fundamental difference appears in the *spectrum* of an infinite-dimensional operator. Although spectral analysis has not yet been applied to global illumination, it illustrates another way in which infinite-dimensional operators differ from matrices. For any scalar value λ, we define the *resolvent operator* of $\mathbf{T}$ by

$$\mathbf{R}_\lambda(\mathbf{T}) \;\equiv\; (\,\mathbf{T} - \lambda\mathbf{I}\,)^{-1}. \tag{11}$$

The *spectrum* of $\mathbf{T}$ consists of all the points λ in the complex plane for which $\mathbf{R}_\lambda$ fails to be a bounded linear operator [14]. For instance, $\mathbf{T}-\lambda\mathbf{I}$ may have a non-trivial *nullspace*, which means that there exists a vector $x \neq 0$ such that

$$(\mathbf{T} - \lambda\mathbf{I})\; x = 0. \tag{12}$$

In this case the resolvent operator $\mathbf{R}_\lambda$ does not exist since $\mathbf{T}-\lambda\mathbf{I}$ cannot be inverted. The value λ is called an *eigenvalue* of $\mathbf{T}$ and the corresponding x is called an *eigenvector*, or an *eigenfunction* when x is a function. In a finite-dimensional space the eigenvalues are the only values at which the resolvent operator $\mathbf{R}_\lambda$ goes awry. In infinite-dimensional spaces, however, there are other possibilities. For example, the resolvent operator may exist but fail to be bounded. The collection of all values at which this occurs is called the *continuous spectrum* of $\mathbf{T}$. As we shall see, many operators associated with global illumination cannot exhibit this feature, which gives them a greater resemblance to matrices.

4.2 Linear Operators for Global Illumination

The classical governing equation can be concisely expressed as a linear operator equation [12, 11, 6, 1]. We shall use a very general formulation [1] similar to that proposed by Gershbein et al. [8]. We first define the *local reflection operator* $\mathbf{K}$ by

$$(\mathbf{K}h)(\mathbf{r},\mathbf{u}) \equiv \int_{\Omega_i} k(\mathbf{r};\mathbf{u}' \to \mathbf{u})\; h(\mathbf{r},\mathbf{u}')\; d\mu(\mathbf{u}'). \tag{13}$$

which accounts for the scattering of incident radiant energy. Here h is a *field radiance function*, that is, a radiance function corresponding to all incident light, which is distinct

from irradiance. The **K** operator maps the field radiance function h to the corresponding surface radiance function after one *local* reflection. The symbol for this operator is a mnemonic for "kernel".

Next, we define the *field radiance operator* **G** that expresses the field radiance at each point of $\mathcal{M}$ in terms of the surface radiance of surrounding visible surfaces. Thus,

$$(\mathbf{G}h)(\mathbf{r},\mathbf{u}) \equiv \begin{cases} h(\mathbf{p}(\mathbf{r},-\mathbf{u}),\mathbf{u}) & \text{when } \nu(\mathbf{r},\mathbf{u}) < \infty \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where h is a surface radiance function. The **G** operator expresses the transport of radiant energy from surface to surface as a linear transformation on the space of surface radiance functions. The symbol chosen for this operator is intended as a mnemonic for "global" or "geometry". Defining **G** in this way allows us to factor out the implicit function $\mathbf{r}'(\mathbf{r},\mathbf{u})$ from the integral in equation (1). In terms of these operators, we may write

$$f = f_0 + \mathbf{KG}\, f. \tag{15}$$

This formulation retains the full generality of equation (1), but is frequently more amenable to analysis. Since the operators express the process of energy transfer, they are subject to various thermodynamic constraints. We next show how these constraints allow us to compute the operator norms.

4.3 The L_1-norm of K

We compute the operator norm $||\,\mathbf{K}\,||_1$ in two steps: first, we produce an upper bound, and then show that the bound is attained. We begin by considering the function norm $||\,\mathbf{K}f\,||_1$ for an arbitrary function $f \in L_1(\mu \times m)$ and obtain a bound in terms of $||\,f\,||_1$ and a new expression, which will be shown to be $||\,\mathbf{K}\,||_1$. Since all quantities are positive, absolute value signs may be dropped. If f is a field radiance function, then

$$\begin{aligned} ||\,\mathbf{K}f\,||_1 &= \int_{\mathcal{M}} \int_{\Omega_o} \int_{\Omega_i} k(\mathbf{r};\mathbf{u}' \to \mathbf{u})\, f(\mathbf{r},\mathbf{u}')\; d\mu(\mathbf{u}')\; d\mu(\mathbf{u})\; dm(\mathbf{r}) \\ &= \int_{\mathcal{M}} \int_{\Omega_i} \int_{\Omega_o} k(\mathbf{r};\mathbf{u}' \to \mathbf{u})\, f(\mathbf{r},\mathbf{u}')\; d\mu(\mathbf{u})\; d\mu(\mathbf{u}')\; dm(\mathbf{r}) \\ &= \int_{\mathcal{M}} \int_{\Omega_i} f(\mathbf{r},\mathbf{u}') \left[\int_{\Omega_o} k(\mathbf{r};\mathbf{u}' \to \mathbf{u})\; d\mu(\mathbf{u}) \right] d\mu(\mathbf{u}')\; dm(\mathbf{r}), \end{aligned}$$

where the first equality follows from the definitions of **K** and the L_1-norm. The second equality holds by Fubini's theorem [20, p. 164], which states that the order of integration can be changed when the integral exists and is finite. In the final equality, we have isolated the kernel of the integral operator **K** to the extent possible. Next, we define the constant ω by

$$\omega \;\equiv\; \operatorname*{ess\,sup}_{\mathbf{r}\in\mathcal{M}} \operatorname*{ess\,sup}_{\mathbf{u}'\in\Omega_i} \int_{\Omega_o} k(\mathbf{r};\mathbf{u}' \to \mathbf{u})\, d\mu(\mathbf{u}), \tag{16}$$

which is independent of the function f. Since this quantity bounds the effect of the bracketed expression on the outer double integral above, we have

$$|| \mathbf{K} f ||_1 \;\leq\; \omega \int_{\mathcal{M}} \int_{\Omega_i} f(\mathbf{r}, \mathbf{u}') \; d\mu(\mathbf{u}') \; dm(\mathbf{r}) \;=\; \omega \, || f ||_1 , \tag{17}$$

which shows that ω is an *upper bound* on $|| \mathbf{K} ||_1$. To show that it is a lower bound as well we must demonstrate that this bound is either attained by some function f, or approached from below by a sequence of functions. We can accomplish the latter by a sequence of beams that approach perfect collimation about the incident direction $\mathbf{u}'$ in which

$$\int_{\Omega_o} k(\mathbf{r}; \mathbf{u}' \rightarrow \mathbf{u}) \, d\mu(\mathbf{u}) \tag{18}$$

is maximal. If this maximum is not attained, but rather approached as the incident beam approaches grazing, for example, then we let the sequence $f_1, f_2, \ldots$ approach perfect collimation while simultaneously approaching grazing. From this it follows that ω is also a *lower bound* on $|| \mathbf{K} ||_1$. Therefore,

$$|| \mathbf{K} ||_1 = \operatorname*{ess\,sup}_{\mathbf{r} \in \mathcal{M}} \; \operatorname*{ess\,sup}_{\mathbf{u}' \in \Omega_i} \int_{\Omega_o} k(\mathbf{r}; \mathbf{u}' \rightarrow \mathbf{u}) \, d\mu(\mathbf{u}) . \tag{19}$$

But the integral in this expression is the directional-hemispherical reflectivity at $\mathbf{r}$ with respect to the incoming direction $\mathbf{u}'$. This quantity is necessarily positive and bounded above by one by the first law of thermodynamics, conservation of energy. Hence, $\omega \leq 1$.

4.4 The L_∞-norm of K

Computing the L_∞-norm proceeds similarly. For any $f \in L_\infty$ we have

$$\begin{aligned} || \mathbf{K} f ||_\infty &= \operatorname*{ess\,sup}_{\mathbf{r} \in \mathcal{M}} \; \operatorname*{ess\,sup}_{\mathbf{u} \in \Omega_o} \int_{\Omega_i} k(\mathbf{r}; \mathbf{u}' \rightarrow \mathbf{u}) \, f(\mathbf{r}, \mathbf{u}') \; d\mu(\mathbf{u}') \\ &\leq \left[\operatorname*{ess\,sup}_{\mathbf{r} \in \mathcal{M}} \; \operatorname*{ess\,sup}_{\mathbf{u} \in \Omega_o} \int_{\Omega_i} k(\mathbf{r}; \mathbf{u}' \rightarrow \mathbf{u}) \; d\mu(\mathbf{u}') \right] || f ||_\infty , \end{aligned}$$

so the expression in brackets is an upper bound on $|| \mathbf{K} ||_\infty$. Again, this bound can be approached from below by considering a sequence of increasingly collimated beams about a direction of maximal reflectance. Therefore, we have

$$|| \mathbf{K} ||_\infty = \operatorname*{ess\,sup}_{\mathbf{r} \in \mathcal{M}} \; \operatorname*{ess\,sup}_{\mathbf{u} \in \Omega_o} \int_{\Omega_i} k(\mathbf{r}; \mathbf{u}' \rightarrow \mathbf{u}) \, d\mu(\mathbf{u}'), \tag{20}$$

which differs from the L_1 norm by exchanging the arguments of the kernel. The integral in this expression is the hemispherical-directional reflectivity at $\mathbf{r}$ with respect to the outgoing direction $\mathbf{u}$. This quantity is necessarily equal to $|| \mathbf{K} ||_1$ because of the reciprocity relation $k(\mathbf{r}; \mathbf{u}' \rightarrow \mathbf{u}) = k(\mathbf{r}; \mathbf{u} \rightarrow \mathbf{u}')$, which is a consequence of the second law of thermodynamics.

4.5 Other Bounds and their Consequences

The L_1 and L_∞ norms for $\mathbf{K}$ extend immediately to all L_p-norms by means of a useful inequality for integral operators, which states that

$$|| \mathbf{K} ||_p \;\leq\; || \mathbf{K} ||_\infty^{(p-1)/p} \; || \mathbf{K} ||_1^{1/p}. \tag{21}$$

Thus, $|| \mathbf{K} ||_p \leq \omega$. See Kato [13, p. 144] or Dunford and Schwartz [7, p. 518].

The $\mathbf{G}$ operator is the easiest to bound since it merely redistributes radiance along straight lines. It follows that $|| \mathbf{G} ||_p = 1$ if any two subsets of $\mathcal{M}$ with non-zero areas can see each other, and $|| \mathbf{G} ||_p = 0$ otherwise. Thus, $|| \mathbf{KG} ||_p \leq || \mathbf{K} ||_p || \mathbf{G} ||_p \leq \omega$. It is interesting to note that if the L_p-norms were based upon the measure σ rather than μ, the $\mathbf{G}$ operator would be unbounded for most geometries.

Given bounds on both $\mathbf{K}$ and $\mathbf{G}$, we next consider the operator

$$\mathbf{M} \;\equiv\; \mathbf{I} - \mathbf{KG}. \tag{22}$$

If $|| \mathbf{K} ||_p < 1$, it follows that $\mathbf{M}^{-1}$ can be expanded into the Neumann series[3]

$$\mathbf{M}^{-1} = \mathbf{I} + \mathbf{KG} + (\mathbf{KG})^2 + \cdots, \tag{23}$$

which converges with respect to all L_p-norms. Since $|| (\mathbf{KG})^n ||_p \leq \omega^n$, we have

$$\left|\left| \mathbf{M}^{-1} \right|\right|_p \leq 1 + \omega + \omega^2 + \ldots = \frac{1}{1-\omega} \tag{24}$$

for all $1 \leq p \leq \infty$. Note that this fails when $\omega = 1$, which is true when perfect reflection occurs at grazing or because of total internal reflection. We therefore ignore these effects and assume that $\omega < 1$. From this bound it follows that $L_p(\mu \times m)$ is *closed* under global illumination. To see this, suppose that $f = \mathbf{M}^{-1} f_0$, where $f_0 \in L_p$. Then the bound on $\mathbf{M}^{-1}$ implies that $\| f \|_p \leq \| \mathbf{M}^{-1} \|_p \| f_0 \|_p < \infty$. Therefore the equilibrium solution f is also in L_p, so we have $\mathbf{M}^{-1} : L_p \to L_p$ for all $1 \leq p \leq \infty$.

Another conclusion that we can draw is that a surface radiance function over $\mathcal{M} \times \mathcal{S}^2$ will decrease in *all* L_p-norms as a result of each interreflection, which follows immediately from the bound $|| \mathbf{KG} ||_p < 1$.

4.6 Adjoint Operators

The natural generalization of "transpose" for linear a operator is the *adjoint*, which has been applied to importance-driven global illumination [24, 2]. While adjoints can be defined in a general Banach space, the symmetry between a Hilbert space and its dual leads to a simple definition of the adjoint. Given a linear operator $\mathbf{M} : \mathcal{H} \to \mathcal{H}$, the so called *Hilbert-adjoint* $\mathbf{M}^*$ is the operator that satisfies

$$\langle \mathbf{M} f \mid g \rangle = \langle f \mid \mathbf{M}^* g \rangle \tag{25}$$

[3] Although this series is frequently attributed to John von Neumann, it is actually due to Carl Gottfried Neumann (1832-1925).

for all f and g in $\mathcal{H}$. In order to define the Hilbert-adjoints of **K** and **G**, we equate surface and field radiance functions so that both operators map $\mathcal{H}$ into itself. To make the necessary identification, we note that the space of field radiance functions is naturally isomorphic to the space of surface radiance functions under the mapping

$$(\mathbf{H}f)\,(\mathbf{r},\mathbf{u}) \equiv f(\mathbf{r},-\mathbf{u})\,. \tag{26}$$

By virtue of this identification, the distinction between surface and field radiance functions can be removed, although it serves as a reminder of the physical process. Treating **K** and **G** as operators that map L_2 into itself, it is easy to see from definition (25) that both are self-adjoint. This fact hinges on associating the cosine factor with the measure μ and not with the kernel k; doing so preserves the symmetry of k and makes the corresponding **K** operator self-adjoint; that is, $\mathbf{K} = \mathbf{K}^*$. It follows that

$$\mathbf{M}^* \;=\; (\mathbf{I}-\mathbf{KG})^* \;=\; \mathbf{I}-\mathbf{GK}.$$

Thus, the same operators **K** and **G** are sufficient to form both **M** and its adjoint. This fact and the connection with the space L_2 are clarified by the operator formulation.

4.7 Compact Operators

As a final topic, we briefly explore the concept of *compactness* [15, p. 40]. A subset S of a Banach space is compact if every infinite sequence of points $x_1, x_2, \ldots$ in S has a limit point in S; that is, some point in S is attained or approached arbitrarily closely (with respect to the norm) infinitely many times by elements of the sequence. In $\mathbb{R}^n$ every closed and bounded subset is compact, but this is not so in infinite dimensions.

A *compact operator* is one that maps bounded sets to compact sets [25, p. 293]. Thus, compact operators remove one of the oddities of infinite-dimensional spaces, which frequently simplifies things. Note that the identity operator **I** is compact if and only if the space it is defined on is finite. An important property of compact operators is that they may be uniformly approximated by degenerate operators; that is, given any compact operator **T** and $\epsilon > 0$, there exists a finite-dimensional operator $\mathbf{T}_n$ such that

$$||\,\mathbf{T}-\mathbf{T}_n\,||_\infty < \epsilon. \tag{27}$$

Compact operators also have a simple *point spectrum*; that is, they have an empty continuous spectrum, as does a finite matrix. As an example of a compact operator, let k be a kernel function such that

$$\int_{\mathcal{M}}\int_{\mathcal{S}^2}\int_{\mathcal{S}^2} |\,k(\mathbf{r},\mathbf{u}'\rightarrow\mathbf{u})\,|^2\,d\mu(\mathbf{u}')\,d\mu(\mathbf{u})\,dm(\mathbf{r}) \;<\; \infty, \tag{28}$$

which means that $k \in L_2(\mu\times\mu\times m)$. Then k is called a *Hilbert-Schmidt kernel* and the corresponding integral operator acting on L_2 is compact [10, p. 17]. Thus, **K** is compact when the reflectance function is a Hilbert-Schmidt function.

As an example of a linear operator that appears in global illumination that is not compact, consider the *specular reflection operator* **S**, defined by

$$(\mathbf{S}f)(\mathbf{r},\mathbf{u}) \;\equiv\; k_s(\mathbf{r};\mathbf{u}'\rightarrow\mathbf{u})\,f(\mathbf{r},\mathbf{u}'), \tag{29}$$

where k_s is the specular kernel, a dimensionless function that is bounded away from one, and $\mathbf{u}'$ denotes the mirror reflection of $\mathbf{u}$. It is easy to see that $||\,\mathbf{S}\,||_p \leq 1$. The operator $\mathbf{S}$ can be factored into a product of a reflection operator (similar to $\mathbf{H}$) and a multiplication operator (similar to $\mathbf{M}_g$). But multiplication operators are not compact, and this property is inherited by $\mathbf{S}$.

To see how these facts affect the process of global illumination, consider the two-pass solution process proposed by Sillion and Puech [23]. We replace equation (15) by $f = f_0 + (\mathbf{K} + \mathbf{S})\mathbf{G}f$, which includes specular reflection, and write it as

$$(\mathbf{I} - \mathbf{SG})f = f_0 + \mathbf{KG}f.$$

Since $||\,\mathbf{S}\,|| < 1$, it follows that $\widehat{\mathbf{S}} \equiv (\mathbf{I} - \mathbf{SG})^{-1}$ exists. Letting $g \equiv (\mathbf{I} - \mathbf{SG})f$ we obtain the two-stage method

$$\begin{aligned} g &= f_0 + \mathbf{KG}\widehat{\mathbf{S}}\,g \\ f &= \widehat{\mathbf{S}}\,g\,, \end{aligned}$$

where the second equation accounts for specular transport chains reaching the eye. Then solving for g in the "first pass" is equivalent to inverting a new operator that includes specular reflections, $\widehat{\mathbf{M}} \equiv \mathbf{I} - \mathbf{KG}\widehat{\mathbf{S}}$. The Neumann series is

$$\widehat{\mathbf{M}}^{-1} = \mathbf{I} + \mathbf{KG}\widehat{\mathbf{S}} + (\mathbf{KG}\widehat{\mathbf{S}})^2 + \cdots. \tag{30}$$

The products of a compact operators with a bounded operator is compact, and a convergent sequence of compact operators is again a compact operator. It follows that $\widehat{\mathbf{M}}^{-1} - \mathbf{I}$ is compact, so the function $g - f_0$, which accounts for all transport chains ending in a non-specular scatter, is the image of a compact operator. Hence, the mapping from f_0 to $g - f_0$ can be approximated to any degree by finite-dimensional operators. This is not true, however, of the mapping from g to f, which does not correspond to a compact operator. This observation provides an abstract view of the difference between specular and non-specular transfers.

5 Further Reading

Functional analysis is a very rich subject that can provide a better theoretical foundation for global illumination. For more complete discussions of the concepts touched upon in this paper, see Glassner [9], Kreyszig [14], Taylor and Lay [25], or Boccara [4]. See Luenberger [15] for a very readable account of Hilbert spaces. For more advanced treatments, see Rudin [19], Kato [13], or Dunford and Schwartz [7].

Acknowledgments

Many thanks to Peter Shirley and Erin Shaw for their helpful comments. This research was supported by the NSF/ARPA Science and Technology Center for Computer Graphics and Scientific Visualization (ASC-8920219) and the Hewlett-Packard Corporation.

References

1. James Arvo, Kenneth Torrance, and Brian Smits. A framework for the analysis of error in global illumination algorithms. In *Computer Graphics* Proceedings, Annual Conference Series, ACM SIGGRAPH, pages 75–84, 1994.
2. Larry Aupperle and Pat Hanrahan. Importance and discrete three point transport. In *Proceedings of Eurographics 93*, pages 85–94, 1993.
3. Marcel Berger. *Geometry, Volume II*. Springer-Verlag, New York, 1987. Translated by M. Cole and S. Levy.
4. Nino Boccara. *Functional Analysis, an Introduction for Physicists*. Academic Press, New York, 1990.
5. S. Chandrasekar. *Radiative Transfer*. Dover Publications, New York, 1960.
6. Michael F. Cohen and John R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press, New York, 1993.
7. Nelson Dunford and Jacob T. Schwartz. *Linear Operators. Part I: General Theory*. John Wiley & Sons, New York, 1967.
8. Reid Gershbein, Peter Schröder, and Pat Hanrahan. Textures and radiosity: Controlling emission and reflection with texture maps. In *Computer Graphics* Proceedings, Annual Conference Series, ACM SIGGRAPH, pages 51–58, 1994.
9. Andew Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann, New York, 1995.
10. Paul R. Halmos and V. S. Sunder. *Bounded Integral Operators on L^2 Spaces*. Springer-Verlag, New York, 1978.
11. Paul S. Heckbert. *Simulating Global Illumination Using Adaptive Meshing*. PhD thesis, University of California, Berkeley, June 1991.
12. James T. Kajiya. The rendering equation. *Computer Graphics*, 20(4):143–150, August 1986.
13. Tosio Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, 1966.
14. Erwin Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, New York, 1978.
15. David G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.
16. Parry Moon. *The Scientific Basis of Illuminating Engineering*. McGraw-Hill, New York, 1936.
17. G. L. Polyak. Radiative transfer between surfaces of arbitrary spatial distribution of reflection. In *Convective and Radiative Heat Transfer*. Publishing House of the Academy of Sciences of the USSR, Moscow, 1960.
18. J. R. Retherford. *Hilbert Space: Compact Operators and the Trace Theorem*. Cambridge University Press, New York, 1993.
19. Walter Rudin. *Functional Analysis*. McGraw-Hill, New York, 1973.
20. Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, second edition, 1974.
21. Irving E. Segal and Ray A. Kunze. *Integrals and Operators*. Springer-Verlag, New York, second edition, 1978.
22. Robert Siegel and John R. Howell. *Thermal Radiation Heat Transfer*. Hemisphere Publishing Corp., New York, second edition, 1981.
23. François Sillion and Claude Puech. A general two-pass method integrating specular and diffuse reflection. *Computer Graphics*, 23(4), August 1989.
24. Brian Smits, James Arvo, and David Salesin. An importance-driven radiosity algorithm. *Computer Graphics*, 26(4):273–282, July 1992.
25. Angus E. Taylor and David C. Lay. *Introduction to Functional Analysis*. John Wiley & Sons, New York, second edition, 1980.
26. Harold Zatz. Galerkin radiosity: A higher order solution method for global illumination. In *Computer Graphics* Proceedings, Annual Conference Series, ACM SIGGRAPH, pages 213–220, 1993.