Finite Element Methods for Radiosity*

Paul S. Heckbert
Carnegie Mellon University

Global Illumination Course
SIGGRAPH '93

There are two general methods for solving the integral equation governing global illumination: finite element methods and Monte Carlo methods. The former approach yields radiosity methods and the latter approach yields ray tracing methods, generally speaking.

Finite element methods are most appropriate for scenes containing only diffuse surfaces, since the radiance function is then a function of only two variables (surface position), and it can be easily discretized. Monte Carlo methods are usually used for general scenes, where the radiance function is a function of four variables (two for surface position and two for direction), and finite element methods would be too cumbersome.

In this chapter we consider finite element methods for diffuse scenes. The phenomenon of global illumination in such scenes is called either *radiosity* or *diffuse interreflection*. Much of the material in this chapter is the application of generic techniques for integral equations and finite element methods, discussed earlier, to the particular problems of radiosity.

We first review numerical solution methods for solving generic integral equations, then we apply those methods to the integral equations governing global illumination in diffuse scenes, and we show that a combination of the simplest finite element techniques yields the radiosity method well known in thermal radiation and computer graphics.

1 Numerical Solution of Integral Equations

Suppose we want to solve the Fredholm integral equation of the second kind,

$$b = e + \mathcal{K}b$$

where $\mathcal{K}$ is the integral operator

$$(\mathcal{K}b)(x) = \int_\Gamma dt\, \kappa(x,t) b(t)$$

The function $e(x)$, the kernel $\kappa(x,t)$, and the domain Γ are known, but $b(x)$ is unknown.

The projection method for the solution of integral equations approximates the exact solution function, which lies in an infinite-dimensional function space, by projecting it to

*Appeared in Global Illumination course notes, SIGGRAPH '93, Anaheim, Aug. 1993

a finite-dimensional function space. As discussed in earlier chapters, the exact solution function $b(x)$ is approximated by a linear combination of basis functions:

$$\hat{b}(x) = \sum_{i=1}^{n} b_i W_i(x) \tag{1}$$

where b_i are the unknown coefficients and W_i are the chosen basis functions.

We would like to find a solution that minimizes the error $\hat{b} - b$, but since the exact solution b is unknown, this is impossible. Instead, we choose $\hat{b}$ so that $\hat{b} \approx e + \mathcal{K}\hat{b}$ is nearly satisfied; the "best" solution in this function space is defined to be the one that minimizes the *residual*

$$r(x) = \hat{b}(x) - (\mathcal{K}\hat{b})(x) - e(x)$$

The minimum can be defined in several ways. The *collocation* method constrains the residual to be zero at a set of points, while the *Galerkin* method constrains the residual function to be orthogonal to each of the basis functions. If n constraints are chosen, then either of these projection methods reduces the integral equation problem to a system of n equations in n unknowns. Since the integral operator $\mathcal{K}$ is linear, this will be a linear system of equations.

1.1 Collocation Method

The collocation method constrains the approximation by requiring that the residual be zero at n collocation points x'_i. For 1-D constant elements, the collocation points are often chosen to be the midpoints of the elements, while for 1-D linear elements, the collocation points are usually the element endpoints. The constraint for each i has the form $r(x'_i) = 0$, so

$$\begin{aligned}
0 &= r(x'_i) \\
&= \hat{b}(x'_i) - (\mathcal{K}\hat{b})(x'_i) - e(x'_i) \\
&= \sum_{j=1}^{n} b_j \Big(W_j(x'_i) - (\mathcal{K}W_j)(x'_i) \Big) - e(x'_i) \\
&= \sum_{j=1}^{n} b_j (M_{ij} - K_{ij}) - e_i
\end{aligned}$$

where

$$\begin{aligned}
M_{ij} &= W_j(x'_i) \\
K_{ij} &= (\mathcal{K}W_j)(x'_i) = \int_\Gamma dt\, \kappa(x'_i, t) W_j(t) \\
e_i &= e(x'_i)
\end{aligned}$$

Collocation thus results in the linear system of equations:

$$(\mathbf{M} - \mathbf{K})\mathbf{b} = \mathbf{e} \tag{2}$$

where $\mathbf{M}$ and $\mathbf{K}$ are $n \times n$ matrices and $\mathbf{b}$ and $\mathbf{e}$ are n-element column vectors. (Boldface denotes matrices and vectors.)

The matrix $\mathbf{M}$ is called the *mass matrix* or *stiffness matrix*. When constant elements are used with collocation points in the interior of each element, or when linear elements are used with a collocation point at each element endpoint, the mass matrix will be the identity: $\mathbf{M} = \mathbf{I}$. For broader basis functions, the more limited the support of the basis functions, the more nearly diagonal the mass matrix will be.

1.2 Galerkin Method

The Galerkin method constrains the approximation by requiring that the residual function be orthogonal to each of the n basis functions. We define the inner product of two real functions f and g over the domain Γ as

$$(f,g) = \int_\Gamma dx\, f(x)g(x)$$

Using this notation, the Galerkin orthogonality constraint for each i is:

$$\begin{aligned} 0 &= (W_i, r) \\ &= (W_i, b) - (W_i, \mathcal{K}b) - (W_i, e) \\ &= \sum_{j=1}^{n} b_j (W_i, W_j) - \sum_{j=1}^{n} b_j (W_i, \mathcal{K}W_j) - (W_i, e) \\ &= \sum_{j=1}^{n} b_j (M_{ij} - K_{ij}) - e_i \end{aligned}$$

with mass matrix

$$M_{ij} = (W_i, W_j) = \int_\Gamma dx\, W_i(x) W_j(x)$$

discretized kernel

$$K_{ij} = (W_i, \mathcal{K}W_j) = \int_\Gamma dx\, W_i(x) \int_\Gamma dt\, \kappa(x,t) W_j(t)$$

and homogeneous vector

$$e_i = (W_i, e) = \int_\Gamma dx\, W_i(x) e(x)$$

As with collocation, Galerkin transforms the integral equation into a linear system of equations:

$$(\mathbf{M} - \mathbf{K})\mathbf{b} = \mathbf{e} \tag{3}$$

Computing the matrix $\mathbf{K}$ with Galerkin methods requires double integration of the kernel, but with collocation methods, only single integration is required. In either case, these integrals often cannot be done analytically, and must be done numerically. The mass matrix is an identity iff the basis functions are orthonormal.

1.3 Comparison of Numerical Solution Techniques

Collocation makes use of the kernel of the integral operator only along constant-x slices of xt space, whereas Galerkin makes use of the entire $\Gamma \times \Gamma$ domain of xt space. Thus, from an intuitive standpoint, Galerkin extracts more information from the kernel than collocation does, allowing it to approximate the integral operator more robustly.

2 Radiosity as a Finite Element Method

Solving radiosity problems using a finite element approach involves several steps [Heckbert-Winget91]. Figure 1 shows some of the available options at each step: one much choose

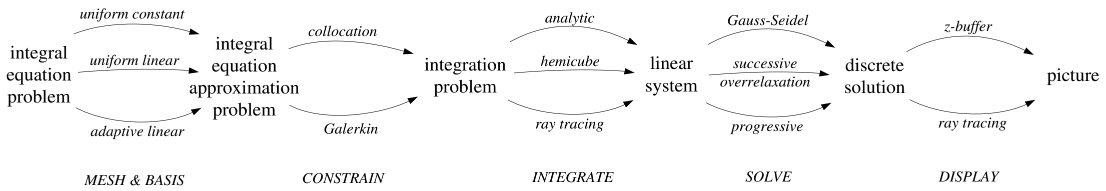

Figure 1: *Steps of finite element simulation of radiosity.*

a mesh and basis functions, constraint and integration methods to reduce the problem to a linear system of equations, a system solution method, and a display algorithm.

Radiosity obeys the integral equation:

$$b(\mathbf{x}) = e(\mathbf{x}) + \int_\Gamma d\mathbf{x}'\, \kappa(\mathbf{x}, \mathbf{x}') b(\mathbf{x}')$$

where $b(\mathbf{x})$ denotes the radiosity at 3-D point $\mathbf{x}$, $e(\mathbf{x})$ is radiant emitted flux density, and Γ is the two-dimensional domain of all surfaces in the scene. Radiosity is the sum of emitted and reflected radiation over the hemisphere of directions. It has units of power/area. For a diffuse surface, radiosity equals π times the outgoing radiance. A more standard term for "radiosity" is "radiant exitance", but we will use the former for historical reasons.

The kernel of the integral equation is

$$\kappa(\mathbf{x}, \mathbf{x}') = \rho_d(\mathbf{x}) \frac{\cos\theta \cos\theta'}{\pi r^2}\, v$$

where $\mathbf{x}'$ is another surface point, $\rho_d(\mathbf{x})$ is the diffuse reflectance (unitless), θ is the polar angle at $\mathbf{x}$, θ' is the polar angle at $\mathbf{x}'$, r is the distance between $\mathbf{x}$ and $\mathbf{x}'$, and v, the visibility function, is 1 if points $\mathbf{x}$ and $\mathbf{x}'$ are inter-visible; 0 if occluded. θ, θ', v, and r are functions of $\mathbf{x}$ and $\mathbf{x}'$.

If we solve this equation using the collocation method for constraint, the following system of equations results:

$$(\mathbf{M} - \mathbf{K})\mathbf{b} = \mathbf{e}$$

where $\mathbf{M}$ and $\mathbf{K}$ are $n \times n$ matrices and $\mathbf{b}$ and $\mathbf{e}$ are n-element column vectors:

$$\begin{aligned} M_{ij} &= W_j(\mathbf{x}_i) \\ K_{ij} &= \int_\Gamma d\mathbf{x}'\, \kappa(\mathbf{x}_i, \mathbf{x}') W_j(\mathbf{x}') \\ e_i &= e(\mathbf{x}_i) \end{aligned}$$

where $\mathbf{x}_i$, the collocation points, are typically the centers or vertices of the elements.

The radiosity method, which was developed in the field of thermal radiation [Hottel54,Sparrow63] and later adapted and optimized for computer graphics [Cohen-Greenberg85,Nishita-Nakamae85], can be derived and explained quite clearly in the context of these numerical methods. The radiosity method typically assumes, for the interreflection simulation, that each polygonal element has a constant radiosity. This corresponds to a mesh with constant elements.

Next, the radiosity method typically computes point-to-area *form factors* from the centers of the elements. (Variants of this method are possible, which compute form factors at polygon vertices instead of the centers, or which compute area-to-area form factors, but point-to-area form factors are the most commonly used.)

One then constructs an $n \times n$ matrix $A_{ij} = I_{ij} - \rho_i F_{ij}$ and solves the system of equations $\mathbf{Ab} = \mathbf{e}$ for the radiosity vector $\mathbf{b}$. But notice that this is precisely the collocation method! In the collocation method we construct the matrix $\mathbf{K}$ which is related to the form factors by $K_{ij} = \rho_i F_{ij}$. If constant elements are used, then $\mathbf{M} = \mathbf{I}$. We then solve the system of equations $(\mathbf{I} - \mathbf{K})\mathbf{b} = \mathbf{e}$. This is the same system of equations as is used in the radiosity method, where $\mathbf{A} = \mathbf{I} - \mathbf{K}$.

The most difficult factor in these integrations is the visibility function v, but this can be evaluated using a hemicube [Cohen-Greenberg85] or ray tracing [Wallace et al. 89]. It is possible to compute certain form factors analytically in two-dimensional [Heckbert91] and three-dimensional [Schröder-Hanrahan93] scenes.

A variety of system solution methods can be used, including generic methods such as Gauss-Seidel iteration, its faster variant successive overrelaxation [Golub-Van Loan89], or specialized methods such as progressive radiosity [Cohen et al. 88]. The latter method can be viewed as an iterative method for solving linear systems that computes one column of the matrix at a time. Display can be done using z-buffer algorithms or ray tracing, similar to integration.

3 Radiosity Research

Recent radiosity research can be understood in the context of the approximation methods described here. We attempt to point the reader to a number of the better recent articles on these topics.

The smoothness of the kernel in the integral equation has been exploited by adaptive, quadtree-like subdivision in xt space, instead of discretizing in a grid [Hanrahan et al. 91,Gortler et al. 93,Schröder et al. 93]. For a problem with m input polygons and n elements (typically, $m \ll n$), solving the system of equations with Gaussian elimination (seldom done in practice) would require $O(n^3)$ work, solving with k steps of Successive Overrelaxation takes $O(kn^2)$ work, solving with k steps of progressive radiosity requires $O(kn)$ work (but suffers from poor accuracy unless k is large), and hierarchical radiosity and its related wavelet methods requires $O(m^2 + n)$ work. The latter complexity figure indicates that these algorithms are best suited to scenes with a small number of large input polygons, finely subdivided. Work needs to be done to optimize these algorithms for complex scenes (m on the order of hundreds of thousands or millions of polygons).

The radiosity method has been generalized to non-diffuse scenes using spherical harmonics as an approximation to directional distributions [Sillion et al. 91]. Higher degree elements [Heckbert-Winget91,Max-Allison92,Salesin et al. 92,Troutman-Max93], Galerkin methods [Heckbert-Winget91,Zatz93], and improved integration techniques [Schröder93,Pietrek93] have been used to improve the accuracy of an approximation.

The following two chapters discuss radiosity in two and three dimensional scenes, respectively, with qualitative analysis of discontinuities caused by occlusion, and the development of better mesh generation algorithms. Similar methods have also been published recently by others in the areas of mesh generation [Campbell91,Lischinski et al. 92,Lischinski et al. 93] and in visibility [Teller92,Teller-Hanrahan93].

References

[Campbell91] A. T. Campbell, III. *Modeling Global Diffuse Illumination for Image Synthesis.* PhD thesis, CS Dept, University of Texas at Austin, Dec. 1991. Tech. Report TR-91-39.

[Cohen et al. 88] Michael F. Cohen, Shenchang Eric Chen, John R. Wallace, and Donald P. Greenberg. A progressive refinement approach to fast radiosity image generation. *Computer Graphics (SIGGRAPH '88 Proceedings)*, 22(4):75–84, Aug. 1988.

[Cohen-Greenberg85] Michael F. Cohen and Donald P. Greenberg. The hemi-cube: A radiosity solution for complex environments. *Computer Graphics (SIGGRAPH '85 Proceedings)*, 19(3):31–40, July 1985.

[Golub-Van Loan89] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, MD, 1989.

[Gortler et al. 93] Steven J. Gortler, Peter Schröder, Michael F. Cohen, and Pat Hanrahan. Wavelet radiosity. *Computer Graphics (SIGGRAPH '93 Proceedings)*, August 1993.

[Hanrahan et al. 91] Pat Hanrahan, David Salzman, and Larry Aupperle. A rapid hierarchical radiosity algorithm. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 25(4):197–206, July 1991.

[Heckbert91] Paul S. Heckbert. *Simulating Global Illumination Using Adaptive Meshing.* PhD thesis, CS Division, UC Berkeley, June 1991. Tech. Report UCB/CSD 91/636.

[Heckbert-Winget91] Paul S. Heckbert and James M. Winget. Finite element methods for global illumination. Technical report, CS Division, UC Berkeley, July 1991. UCB/CSD 91/643.

[Hottel54] Hoyt C. Hottel. Radiant heat transmission. In W. H. McAdams, editor, *Heat Transmission, 3rd ed.*, New York, 1954. McGraw-Hill.

[Lischinski et al. 92] Dani Lischinski, Filippo Tampieri, and Donald P. Greenberg. Discontinuity meshing for accurate radiosity. *IEEE Computer Graphics and Applications*, 12(6):25–39, November 1992.

[Lischinski et al. 93] Dani Lischinski, Filippo Tampieri, and Donald P. Greenberg. Combining hierarchical radiosity and discontinuity meshing. *Computer Graphics (SIGGRAPH '93 Proceedings)*, August 1993.

[Max-Allison92] Nelson L. Max and Michael J. Allison. Linear radiosity approximation using vertex-to-vertex form factors. In David Kirk, editor, *Graphics Gems III*, pages 318–323. Academic Press, 1992.

[Nishita-Nakamae85] Tomoyuki Nishita and Eihachiro Nakamae. Continuous tone representation of 3-D objects taking account of shadows and interreflection. *Computer Graphics (SIGGRAPH '85 Proceedings)*, 19(3):23–30, July 1985.

[Pietrek93] Georg Pietrek. Fast calculation of accurate formfactors. In François Sillion and Michael F. Cohen, editors, *Fourth Eurographics Workshop on Rendering*, Paris, June 1993.

[Salesin et al. 92] David Salesin, Dani Lischinski, and Tony DeRose. Reconstructing illumination functions with selected discontinuities. In Alan Chalmers and Derek Paddon, editors, *Third Eurographics Workshop on Rendering*, pages 99–112, Bristol, UK, May 1992.

[Schröder et al. 93] Peter Schröder, Steven J. Gortler, Michael F. Cohen, and Pat Hanrahan. Wavelet projections for radiosity. In François Sillion and Michael F. Cohen, editors, *Fourth Eurographics Workshop on Rendering*, Paris, June 1993.

[Schröder93] Peter Schröder. Numerical integration for radiosity in the presence of singularities. In François Sillion and Michael F. Cohen, editors, *Fourth Eurographics Workshop on Rendering*, Paris, June 1993.

[Schröder-Hanrahan93] Peter Schröder and Pat Hanrahan. On the form factor between two polygons. *Computer Graphics (SIGGRAPH '93 Proceedings)*, August 1993.

[Sillion et al. 91] François X. Sillion, James R. Arvo, Stephen H. Westin, and Donald P. Greenberg. A global illumination solution for general reflectance distributions. *Computer Graphics (SIGGRAPH '91 Proceedings)*, 25(4):187–196, July 1991.

[Sparrow63] Ephraim M. Sparrow. On the calculation of radiant interchange between surfaces. In Warren Ibele, editor, *Modern Developments in Heat Transfer*, New York, 1963. Academic Press.

[Teller92] Seth J. Teller. Computing the antipenumbra of an area light source. *Computer Graphics (SIGGRAPH '92 Proceedings)*, 26(2):139–148, July 1992.

[Teller-Hanrahan93] Seth J. Teller and Pat Hanrahan. Global visibility algorithms for illumination computations. *Computer Graphics (SIGGRAPH '93 Proceedings)*, August 1993.

[Troutman-Max93] Roy Troutman and Nelson Max. Radiosity algorithms using higher-order finite elements. In *Computer Graphics (SIGGRAPH '93 Proceedings)*, pages 209–212, August 1993.

[Wallace et al. 89] John R. Wallace, Kells A. Elmquist, and Eric A. Haines. A ray tracing algorithm for progressive radiosity. *Computer Graphics (SIGGRAPH '89 Proceedings)*, 23(3):315–324, July 1989.

[Zatz93] Harold R. Zatz. Galerkin radiosity: A higher-order solution method for global illumination. *Computer Graphics (SIGGRAPH '93 Proceedings)*, August 1993.