Ian Ashdown
620 Ballantree Road
West Vancouver, BC
Canada V7S 1W3

Tel: (604) 922-6148
Fax: (604) 987-7621
e-mail: ian_ashdown@helios32.com

**IESNA Paper #8**

September 21st, 2001

# Radiative Transfer Networks Revisited

by

Ian Ashdown, P. Eng., LC, FIES
Chief Technology Officer
byHeart Consultants Limited

## Abstract

Radiative flux transfer between Lambertian surfaces can be described in terms of linear resistive networks with voltage sources. In this paper we examine how these "radiative transfer networks" provide a physical interpretation for the eigenvectors of form factor matrices. This leads to a novel approach to photorealistic image synthesis and radiative transfer analysis called *eigenvector radiosity*.

## 1. Introduction

The *radiosity equation*[24,27] is used in computer graphics for photorealistic image synthesis, and in thermal and illumination engineering for radiative and luminous flux transfer calculations. Because it may involve extremely large matrices, considerable efforts have been devoted to finding fast iterative solution methods that do not consume large amounts of memory.

In the 1950's, thermal and illumination engineering researchers noted that the net flow of infrared and visible light between surfaces is analogous to the flow of electrical current in linear resistive networks. This analogy was investigated in depth, but it unfortunately did not lead to useful techniques for solving the radiosity equation.

In this paper we revisit this "radiative transfer (RT) network" analogy. We demonstrate that it offers an interesting and instructive physical interpretation of the symmetric form of the radiosity equation.

Using this interpretation, we find that it explains the physical significance of the eigenvectors of form factor matrices[3,4,6]. This in turn leads to *eigenvector radiosity* for photorealistic image synthesis, and also for thermal and illuminating engineering analysis.

## 2. Radiosity Equation

Given a system of $n$ discrete surfaces $A_i$ with Lambertian reflectance $\rho_i$, irradiance $E_i$, and possibly zero initial radiant exitance $M_{oi}$ for each surface $i$, the radiosity equation:

$$M_i = M_{oi} + \rho_i E_i = M_{oi} + \rho_i \sum_{j=1}^{n} F_{ij} M_j \tag{1}$$

describes the final radiant exitance $M_i$ of each surface $i$, where $F_{ij}$ is the form factor from surface $i$ to surface $j$, and where $E$ and $M$ are assumed to be constant over each surface. Expressed in matrix form, this becomes:

$$\mathbf{M} = \mathbf{M}_o + \mathbf{RE} = \mathbf{M}_o + \mathbf{RFM} \tag{2}$$

where $\mathbf{M}$ is the final radiant exitance vector, $\mathbf{M}_o$ is the initial radiant exitance vector, $\mathbf{E}$ is the irradiance vector, $\mathbf{R}$ is the diagonal reflectance matrix (where $R_{ii} = \rho_i$ and $R_{ij} = 0,\ i \neq j$), and $\mathbf{F}$ is the form factor matrix.

Rearranged slightly, this becomes:

$$\mathbf{M}_o = (\mathbf{I} - \mathbf{RF})\mathbf{M} \tag{3}$$

which is the familiar matrix form of the radiosity equation.

## 3. Radiative Transfer Networks

As shown by Oppenheim[20] and O'Brien[17], the radiative flux transfer between these surfaces can be described in terms of a linear resistive network with one or more voltage sources. For a given surface $A_i$, the net flux transfer away from the surface is given by:

$$\Phi_i^{net} = A_i\left(M_i - E_i\right) \tag{4}$$

which is analogous to current in an electrical circuit.

Substituting $E_i$ from Equation 1 into Equation 4 gives:

$$\Phi_i^{net} = \frac{A_i\left(1-\rho_i\right)}{\rho_i}\left[\frac{M_{oi}}{\left(1-\rho_i\right)} - M_i\right] \tag{5}$$

where from Ohm's Law[25], the term $\left[\frac{M_{oi}}{\left(1-\rho_i\right)} - M_i\right]$ is analogous to voltage, and the term $\frac{A_i\left(1-\rho_i\right)}{\rho_i}$ is analogous to conductance (i.e., the inverse of resistance).

Now to conserve energy in the system, the following relation must hold for each surface $A_i$:

$$\sum_{j=1}^{n} F_{ij} = 1 \tag{6}$$

This means that if the system is not fully enclosed, it must be enclosed in a box whose interior surfaces have zero reflectance and zero initial radiant exitance. This box will act as a sink to maintain the energy balance.

Therefore, by again substituting $E_i$ from Equation 1 into Equation 4, we also have:

$$\Phi_i^{net} = \sum_{j=1}^{n} A_i F_{ij}\left(M_i - M_j\right) \tag{7}$$

which gives:

$$\sum_{j=1}^{n} A_i F_{ij}\left(M_i - M_j\right) = \frac{A_i\left(1-\rho_i\right)}{\rho_i}\left[\frac{M_{oi}}{\left(1-\rho_i\right)} - M_i\right] \tag{8}$$

This is analogous to Kirchoff's Current Law[25], which states that the algebraic sum of the currents entering a node in an electrical circuit is always zero.

This completes the analogy: each surface represents a node in an electrical circuit, with radiative flux (current) flowing between these surfaces (nodes). The amount of flux is determined by the analogies of voltage and conductance.

Figure 1 shows as an example the RT network for an enclosed system defined by three surfaces with finite width and infinite length. The conductances $g_{o1}$, $g_{o2}$, and $g_{o3}$ represent the initial radiant exitances, and have the form:

$$g_{oi} = \frac{A_i\left(1-\rho_i\right)}{\rho_i} \tag{9}$$

while the other conductances represent the interreflections of flux between the surfaces, and have the form (from the reciprocity relationship $A_iF_{ij} = A_jF_{ji}$):

$$g_{ij} = A_iF_{ij} = A_jF_{ji} = g_{ji} \tag{10}$$

for any two surfaces $A_i$ and $A_j$.

The surface exitances $M_i$ are analogous to the circuit node voltages, while the initial surface exitances $M_{oi}$ are analogous to the node voltages $M_{oi} / \rho_i$. If $M_{oi}$ is zero, the node is connected to ground; otherwise it is connected to a voltage source.

All other RT networks are topologically similar to Figure 1. That is, each surface $A_i$ will have associated with it a resistor with conductance $g_{oi}$ that has one terminal connected to either a fixed voltage source if it has an initial radiant exitance $M_{oi}$ or ground. The other terminal will be connected through resistors with conductances $g_{ij}$ to every other surface $A_j$. Where $F_{ij} = 0$, these resistors will have zero conductance – in other words, an open circuit.

The RT network analogy represents an environment solely in terms of radiative flux transfer between its surfaces. If we are interested in simplifying an environment to reduce the computational effort needed to solve its radiosity equation, then it is reasonable to begin with its RT network.

## 4. Symmetric Matrices

From Equation 8 we can derive these simultaneous linear equations[17]:

$$\begin{aligned}
\frac{A_1 M_{o1}}{\rho_1} &= \frac{A_1 M_1}{\rho_1} - F_{12} A_1 M_2 \cdots - F_{1n} A_1 M_n \\
\frac{A_2 M_{o2}}{\rho_2} &= -F_{21} A_2 M_1 + \frac{A_2 M_2}{\rho_2} \cdots - F_{2n} A_2 M_n \\
&\cdots \\
\frac{A_n M_{on}}{\rho_n} &= -F_{N1} A_N M_1 - F_{N2} A_N M_2 \cdots + \frac{A_N M_n}{\rho_n}
\end{aligned} \tag{11}$$

Expressed in matrix form, these become:

$$\mathbf{A}\mathbf{R}^{-1}\mathbf{M}_o = \mathbf{A}\mathbf{R}^{-1}(\mathbf{I} - \mathbf{R}\mathbf{F})\mathbf{M} \tag{12}$$

which is equivalent to Equation 3 (with $\mathbf{A}$ being the diagonal matrix of surface areas). As noted by Nievergelt[16], multiplying both sides of Equation 3 by $\mathbf{A}\mathbf{R}^{-1}$ in Equation 12 replaces exitances with equivalent radiant fluxes that produce the same exitances from the system's Lambertian surfaces.

Neumann[15] and Nievergelt[16] both noted that this operation produces the radiosity matrix $\mathbf{A}\mathbf{R}^{-1}(\mathbf{I} - \mathbf{R}\mathbf{F})$, which is symmetric positive definite.

Trivially rearranging terms in Equation 2, we also have:

$$\mathbf{M} = \mathbf{M}_o + \mathbf{R}\mathbf{A}^{-1}(\mathbf{A}\mathbf{F})\mathbf{M} = \mathbf{M}_o + \mathbf{S}\mathbf{G}\mathbf{M} \tag{13}$$

where with reference to radiative transfer networks, the matrix $\mathbf{G}$ is the symmetric *conductance matrix* for the interconnected conductances $g_{ij}$.

## 5. Spectral Decomposition

Given any square $n \times n$ matrix $\mathbf{B}$, the *spectral decomposition theorem*[5] states that:

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^* = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^* \tag{14}$$

where $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is a matrix of the orthonormalized eigenvectors of $\mathbf{B}$, $\mathbf{D} = diag(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of their associated eigenvalues, and $\mathbf{v}_i \mathbf{v}_i^*$ is the $i^{th}$ eigenprojection of $\mathbf{B}$ corresponding to the eigenvalue $\lambda_i$.

In many cases, the matrix $\mathbf{B}$ can be approximated by a subset of the eigenvectors with the largest magnitude eigenvalues. That is, if the eigenvectors are ordered such that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$, then:

$$\mathbf{B} \approx \mathbf{B}' = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^*, \quad p << n \tag{15}$$

where $\mathbf{B}'$ is rank deficient. (For a symmetric matrix, the complex conjugate vector $\mathbf{v}_i^*$ becomes the transpose vector $\mathbf{v}_i^{\mathrm{T}}$.)

## 6. Approximate Conductance Networks

Applying the spectral theorem to the symmetric radiative transfer conductance matrix $\mathbf{G}$, we have:

$$\mathbf{G} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathrm{T}} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}} \tag{16}$$

The question is whether this spectral decomposition of $\mathbf{G}$ has any physical meaning. If it does, it may offer a useful approach to solving the radiosity equation.

As an example, consider the empty rectangular room shown in Figure 2 with measurements:

| Length | Width | Height |
|---|---|---|
| 5.0 m | 3.0 m | 2.5 m |

and surface reflectances:

| Ceiling | Walls | Floor |
|---|---|---|
| 80 % | 70 % | 20 % |

From the analytic form factors equations for parallel and perpendicular rectangles (e.g., Howell[9]) and symmetry considerations, we have:

$F_{11} = F_{22} = F_{33} = F_{44} = F_{55} = F_{66} = 0.0000$
$F_{12} = F_{13} = F_{62} = F_{63} = 0.1249$
$F_{14} = F_{15} = F_{64} = F_{65} = 0.2145$
$F_{16} = F_{61} = 0.3213$
$F_{21} = F_{26} = F_{31} = F_{36} = 0.2498$
$F_{23} = F_{32} = 0.0800$
$F_{24} = F_{25} = F_{34} = F_{35} = 0.2102$
$F_{41} = F_{46} = F_{51} = F_{56} = 0.2573$
$F_{42} = F_{43} = F_{52} = F_{53} = 0.1261$
$F_{45} = F_{54} = 0.2331$

This gives the symmetric conductance matrix:

$$\mathbf{G} = \begin{bmatrix} 0.000 & 1.873 & 1.873 & 3.216 & 3.216 & 4.819 \\ 1.873 & 0.000 & 0.600 & 1.576 & 1.576 & 1.873 \\ 1.873 & 0.600 & 0.000 & 1.576 & 1.576 & 1.873 \\ 3.216 & 1.576 & 1.576 & 0.000 & 2.913 & 3.216 \\ 3.216 & 1.576 & 1.576 & 2.913 & 0.000 & 3.216 \\ 4.819 & 1.873 & 1.873 & 3.216 & 3.216 & 0.000 \end{bmatrix}$$

which has the eigenvalues:

$$12.401, \quad -4.819, \quad -2.913, \quad -2.639, \quad -1.429, \quad -0.600$$

and their associated eigenvectors:

$$\mathbf{V} = \begin{bmatrix} -0.495 \\ -0.271 \\ -0.271 \\ -0.425 \\ -0.425 \\ -0.495 \end{bmatrix}, \begin{bmatrix} +0.707 \\ +0.000 \\ +0.000 \\ +0.000 \\ +0.000 \\ -0.707 \end{bmatrix}, \begin{bmatrix} +0.000 \\ +0.000 \\ +0.000 \\ -0.707 \\ +0.707 \\ +0.000 \end{bmatrix}, \begin{bmatrix} -0.470 \\ +0.030 \\ +0.030 \\ +0.527 \\ +0.527 \\ -0.470 \end{bmatrix}, \begin{bmatrix} -0.183 \\ +0.652 \\ +0.652 \\ -0.201 \\ -0.201 \\ -0.183 \end{bmatrix}, \begin{bmatrix} +0.000 \\ -0.707 \\ +0.707 \\ +0.000 \\ +0.000 \\ +0.000 \end{bmatrix}$$

Using the five eigenvectors with the largest magnitude eigenvalues to reconstruct $\mathbf{G}'$ according to Equation 15, we have:

$$\mathbf{G}'_5 = \begin{bmatrix} 0.000 & 1.873 & 1.873 & 3.216 & 3.216 & 4.819 \\ 1.873 & \mathbf{0.300} & \mathbf{0.300} & 1.576 & 1.576 & 1.873 \\ 1.873 & \mathbf{0.300} & \mathbf{0.300} & 1.576 & 1.576 & 1.873 \\ 3.216 & 1.576 & 1.576 & 0.000 & 2.913 & 3.216 \\ 3.216 & 1.576 & 1.576 & 2.913 & 0.000 & 3.216 \\ 4.819 & 1.873 & 1.873 & 3.216 & 3.216 & 0.000 \end{bmatrix}$$

where the differences between $\mathbf{G}$ and $\mathbf{G}'$ are highlighted.

We may interpret each eigenvector $\mathbf{v}_i$ and its associated eigenvalue $\lambda_i$ of matrix $\mathbf{G}$ as describing a radiative transfer network with the same topology as that described by $\mathbf{G}$, but with different conductances between its nodes (surfaces). From Equation 16, its conductance matrix is seen to be $\mathbf{G}(\mathbf{v}_i) = \lambda_i \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$.

Knowing that the conductance of $n$ resistors connected in parallel is:

$$g_{total} = \sum_{i=1}^{n} g_i \tag{17}$$

it is therefore evident that Equation 16 describes the superposition of the networks represented by the eigenprojections of $\mathbf{G}$. Moreover, those eigenvectors with the largest magnitude eigenvalues describe the *dominant paths* (i.e., largest conductances) for radiative flux transfer between the surfaces.

There are several points to note here. First, these paths do not indicate the net flow of radiant flux between surfaces – this is dependent on the initial surface exitances. The eigenprojections describe the network itself rather than any network state.

Second, the diagonal elements $g'_{22}$ and $g'_{33}$ have non-zero values. These can be interpreted as loop conductances connecting the nodes to themselves, as shown in Figure 3.

In an electrical network, a loop conductance has no effect – the voltage is the same at both terminals, and so there is no net flow of current through the conductance.

The same argument applies to radiative transfer networks – there can be no net flow of flux through a loop conductance. This is evident from Equation 7, where the contribution of the form factor $F_{ii}$ is cancelled by the null term ($M_i - M_i$).

One more point regarding the approximate conductance matrix $\mathbf{G}'$ requires a different example. The following conductance matrix $\mathbf{G}$ was generated randomly:

$$\mathbf{G} = \begin{bmatrix} 0.000 & 0.451 & 0.043 & 0.027 & 0.312 & 0.012 \\ 0.451 & 0.000 & 0.384 & 0.683 & 0.092 & 0.035 \\ 0.043 & 0.384 & 0.000 & 0.612 & 0.608 & 0.015 \\ 0.027 & 0.683 & 0.612 & 0.000 & 0.016 & 0.190 \\ 0.312 & 0.092 & 0.608 & 0.016 & 0.000 & 0.586 \\ 0.012 & 0.035 & 0.015 & 0.190 & 0.586 & 0.000 \end{bmatrix}$$

and, using the fact that the sum of each column $i$ represents the associated area $a_i$ to obtain:

$$\mathbf{A} = diag\left[0.848,\ \ 1.646,\ \ 1.664,\ \ 1.529,\ \ 1.617,\ \ 0.841\right]$$

the associated form factor matrix $\mathbf{F} = \mathbf{A}^{-1}\mathbf{G}$ is:

$$\mathbf{F} = \begin{bmatrix} 0.000 & 0.532 & 0.051 & 0.032 & 0.368 & 0.015 \\ 0.274 & 0.000 & 0.233 & 0.414 & 0.056 & 0.021 \\ 0.026 & 0.230 & 0.000 & 0.367 & 0.365 & 0.009 \\ 0.017 & 0.446 & 0.400 & 0.000 & 0.010 & 0.124 \\ 0.193 & 0.057 & 0.376 & 0.010 & 0.000 & 0.362 \\ 0.015 & 0.042 & 0.018 & 0.226 & 0.697 & 0.000 \end{bmatrix}$$

Using the four eigenvectors with the largest magnitude eigenvalues to reconstruct $\mathbf{G}'$ according to Equation 15 gives:

$$\mathbf{G}'_4 = \begin{bmatrix} \mathbf{-0.084} & +0.424 & +0.083 & +0.057 & +0.310 & +0.017 \\ +0.424 & +0.004 & +0.367 & +0.706 & +0.077 & +0.069 \\ +0.083 & +0.367 & +0.044 & +0.568 & +0.641 & \mathbf{-0.057} \\ +0.057 & +0.706 & +0.568 & +0.003 & +0.002 & +0.221 \\ +0.310 & +0.077 & +0.641 & +0.002 & +0.016 & +0.551 \\ +0.017 & +0.069 & \mathbf{-0.057} & 0.221 & +0.551 & +0.078 \end{bmatrix}$$

where the highlighted elements represent negative conductances.

In electrical networks, the linear amplifier has the transfer characteristic $i = mv$, where $v$ is the input voltage, $i$ is the output current, and $m$ is a constant. From Ohm's law, the device conductance is $m$. If $m$ is negative, the amplifier inverts the input signal. If in addition the amplifier gain less than unity, it behaves as it were a negative conductance.

Moon[13] described a radiant flux analogue of a linear amplifier with positive $m$. It consists of an opal glass diffuser with lamps behind it that are controlled by a photocell and dimmer. The radiant flux emitted from the diffuser is directly proportional to the incident flux. An equivalent linear amplifier with negative $m$ (representing negative conductance) can easily be constructed by proportionately dimming the lamps in response to increased incident flux.

This of course requires that the linear amplifier be biased such that it produces a constant non-zero flux output in the absence of incident flux. Otherwise, the amplifier will have to produce "negative" light, a physical impossibility.

This situation is not unusual in electrical networks. For example, tunnel diodes have nonlinear transfer characteristics that include negative conductance behavior over a restricted range of positive input voltages. Within this range, the transfer characteristic is

approximately $i = mv + c$, where $m$ is negative and $c$ is a positive constant. By applying a constant bias voltage to the diode, it behaves as if it were a negative conductance resistor over a range of input voltages.

The same situation may apply to an RT network that includes an inverting linear amplifier – the adjoining conductances may sufficiently "bias" the amplifier such that it behaves like a negative conductance over a range of initial surface exitances.

From Equations 16 and 17, it is therefore evident that any approximate conductance matrix $\mathbf{G}'$ can be represented by a radiative transfer network, even if contains loop and negative conductances. Whether this network is physically realizable (i.e., it does not require "negative" light) depends on the network conductances and the initial surface exitances.

Another problem with negative conductances in both electrical and radiative transfer networks is that they can lead to uncontrolled positive feedback where the solution goes to infinity. Whether this situation will occur in an RT network depends on both the initial surface exitances and the relative magnitude of the negative conductances. (In particular, the non-zero diagonal elements may result in a radiosity matrix that is not diagonally dominant.) This is unlikely to occur however where the differences between $\mathbf{G}$ and $\mathbf{G}'$ are small.

## 7. Approximation Accuracy

Knowing that an approximate conductance matrix can be represented by a radiative transfer network does not unfortunately indicate how accurate the approximation will be. While this will depend on the particular network, it is instructive to consider the empty rectangular room example presented in Figure 2.

Suppose that the initial exitance vector is:

$$\mathbf{M}_o = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}^{\mathrm{T}}$$

Solving for $\mathbf{M}$ directly using the equation:

$$\mathbf{M} = \left(\mathbf{I} - \mathbf{R}\mathbf{A}^{-1}\mathbf{G}\right)^{-1}\mathbf{M}_o \tag{18}$$

we have:

$$\mathbf{M} = \begin{bmatrix} 1.2343 & 0.3684 & 0.3684 & 0.3713 & 0.3713 & 0.1296 \end{bmatrix}^{\mathrm{T}}$$

If we construct the approximate conductance $\mathbf{G}'$ using only the two largest magnitude eigenvectors of $\mathbf{G}$, we have:

$$\mathbf{G}'_2 = \begin{bmatrix} 0.6316 & 1.6643 & 1.6643 & 2.6153 & 2.6153 & 5.4514 \\ 1.6643 & 0.9107 & 0.9107 & 1.4311 & 1.4311 & 1.6643 \\ 1.6643 & 0.9107 & 0.9107 & 1.4311 & 1.4311 & 1.6643 \\ 2.6153 & 1.4311 & 1.4311 & 2.2488 & 2.2488 & 2.6153 \\ 2.6153 & 1.4311 & 1.4311 & 2.2488 & 2.2488 & 2.6153 \\ 5.4514 & 1.6643 & 1.6643 & 2.6153 & 2.6153 & 0.6316 \end{bmatrix}$$

which in comparison with $\mathbf{G}$ does look not at all promising. However, substituting $\mathbf{G}'_2$ for $\mathbf{G}$ in Equation 18 yields:

$$\mathbf{M}'_2 = \begin{bmatrix} 1.2431 & 0.3695 & 0.3695 & 0.3484 & 0.3484 & 0.1322 \end{bmatrix}^{\mathrm{T}}$$

and subtracting $\mathbf{M}'$ from $\mathbf{M}$ gives:

$$\mathbf{M} - \mathbf{M}'_2 = \begin{bmatrix} -0.0088 & -0.0011 & -0.0011 & +0.0229 & +0.0229 & -0.0026 \end{bmatrix}^{\mathrm{T}}$$

Somewhat remarkably, the maximum per-element error between $\mathbf{M}$ and $\mathbf{M}'_2$ is less than 6.2%, and the relative 2-norm error between the two solution vectors is only 2.7%. Thus, while the approximate conductance matrix $\mathbf{G}'$ may not resemble $\mathbf{G}$ through visual inspection, it produces almost the same results when used to solve the radiosity equation.

While it difficult to extrapolate these results to more complex environments, they clearly illustrate the potential of approximate conductance networks.

## 8. Eigenvector Radiosity

Approximating the radiosity matrix $\mathbf{G}$ with the rank deficient approximation $\mathbf{G}'$ does not assist in solving the radiosity equation using conventional solvers such as Gauss-Seidel or Southwell iteration[24]. This is because $\mathbf{G}'$ is still an $n \times n$ matrix. However, Equation 3 can be reformulated to take full advantage of $\mathbf{G}'$.

To begin with, Equation 2 can be solved iteratively as:

$$\mathbf{M}^{(r)} = \mathbf{M}_o + \mathbf{RFM}^{(r-1)} \tag{19}$$

which, following Equation 13, can be reformulated as:

$$\mathbf{M}^{(r)} = \mathbf{M}_o + \mathbf{SGM}^{(r-1)} \tag{20}$$

where $\mathbf{S} = \mathbf{RA}^{-1}$, $\mathbf{G} = \mathbf{AF}$, and $\mathbf{M}^{(0)} = \mathbf{M}_o$.

Because $\mathbf{G}$ is real and symmetric, its eigenvectors $\mathbf{v}_i$ form an orthonormal set spanning an $n$-dimensional space, where $n$ is the matrix order. An arbitrary $n$-dimensional vector $\mathbf{u}$ can always be expressed as a weighted sum of these eigenvectors as:

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{v}_i, \quad i = 1, \ldots, n \tag{21}$$

where the coefficients $a_i$ are given by:

$$a_i = \mathbf{u}^{\mathrm{T}} \mathbf{v}_i, \quad i = 1, \ldots, n \tag{22}$$

The initial exitance vector $\mathbf{M}_o$ can therefore be expressed in terms of $\mathbf{v}_i$ as:

$$\mathbf{M}_o = \sum_{i=1}^{n} \mathbf{M}_o^{\mathrm{T}} \mathbf{v}_i \mathbf{v}_i \tag{23}$$

Substituting this result into Equation 20 gives the first estimate of the exitance vector $\mathbf{M}$ as:

$$\mathbf{M}^{(1)} = \mathbf{M}_o + \mathbf{SGM}_o = \sum_{i=1}^{n} \mathbf{M}_o^{\mathrm{T}} \mathbf{v}_i \mathbf{v}_i + \mathbf{SG} \sum_{i=1}^{n} \mathbf{M}_o^{\mathrm{T}} \mathbf{v}_i \mathbf{v}_i \tag{24}$$

From the definition of an eigenvector as:

$$\mathbf{G}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \ldots, n \tag{25}$$

we also have:

$$\mathbf{G}\sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\mathbf{v}_i = \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\mathbf{G}\mathbf{v}_i = \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{v}_i \tag{26}$$

which gives:

$$\mathbf{M}^{(1)} = \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\mathbf{v}_i + \mathbf{S}\sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{v}_i = \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\mathbf{v}_i + \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i \tag{27}$$

where $\mathbf{w}_i = \mathbf{S}\mathbf{v}_i$.

Using Equation 20 again, the second estimate of the exitance vector $\mathbf{M}$ is:

$$\mathbf{M}^{(2)} = \mathbf{M}_o + \mathbf{S}\mathbf{G}\mathbf{M}^{(1)} \tag{28}$$

where from Equations 21 and 22, we have:

$$\mathbf{M}^{(1)} = \sum_{i=1}^{n}\left(\mathbf{M}^{(1)}\right)^{\mathrm{T}}\mathbf{v}_i\mathbf{v}_i \tag{29}$$

which gives:

$$\mathbf{M}^{(2)} = \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\mathbf{v}_i + \sum_{i=1}^{n}\left(M^{(1)}\right)^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i \tag{30}$$

Using Equation 27 to expand coefficients, we have:

$$\left(\mathbf{M}^{(1)}\right)^{\mathrm{T}}\mathbf{v}_i = \left(\mathbf{M}_o + \mathbf{S}\mathbf{G}\mathbf{M}_o\right)^{\mathrm{T}}\mathbf{v}_i \tag{31}$$

which becomes:

$$\left(\mathbf{M}^{(1)}\right)^{\mathrm{T}}\mathbf{v}_i = \mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i + \sum_{j=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_j\mathbf{v}_i^{\mathrm{T}}\mathbf{w}_j, \quad i = 1,\ldots,n \tag{32}$$

Therefore:

$$\mathbf{M}^{(2)} = \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\mathbf{v}_i + \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i + \sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i^{\mathrm{T}}\mathbf{v}_j\lambda_j\mathbf{w}_j \tag{33}$$

The third estimate is then:

$$\begin{aligned}\mathbf{M}^{(3)} = \mathbf{M}_o + \sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i + \sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i^{\mathrm{T}}\mathbf{v}_j\lambda_j\mathbf{w}_j + \\ \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{M}_o^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i^{\mathrm{T}}\mathbf{v}_j\lambda_j\mathbf{w}_j\mathbf{v}_k\lambda_k\mathbf{w}_k\end{aligned} \tag{34}$$

Applying induction to the first three estimates, it is evident that if:

$$\mathbf{K}^{(0)} = \mathbf{M}_o \tag{35}$$

then:

$$\mathbf{K}^{(r)} = \sum_{i=1}^{n}\left(\mathbf{K}^{(r-1)}\right)^{\mathrm{T}}\mathbf{v}_i\lambda_i\mathbf{w}_i, \quad r \geq 1 \tag{36}$$

and:

$$\mathbf{M}^{(r)} = \mathbf{M}^{(r-1)} + \mathbf{K}^{(r)}, \quad r \geq 1 \tag{37}$$

where to summarize:

$\mathbf{M}_o$ is the initial exitance vector;
$\mathbf{M}^{(r)}$ is the $r^{\text{th}}$ estimate of the exitance vector $\mathbf{M}$;
$\lambda_i$ is the $i^{\text{th}}$ eigenvalue of the conductance matrix $\mathbf{G}$;
$\mathbf{v}_i$ is the $i^{\text{th}}$ eigenvector of the conductance matrix $\mathbf{G}$; and
$\mathbf{w}_i = \mathbf{S}\mathbf{v}_i = \mathbf{R}\mathbf{A}^{-1}\mathbf{v}_i$.

The above reformulation of the radiosity equation in terms of eigenvectors is derived from DiLaura and Franck[6]. Their work was based on Equation 19 and the form factor matrix $\mathbf{F}$ rather than Equation 20 and the conductance matrix $\mathbf{G}$. This limited their derivation to symmetric form factor matrices, which generally occur only in unoccluded environments where all surface elements are the same size. Substituting the conductance matrix $\mathbf{G}$ generalizes eigenvector radiosity to all physically realizable environments.

There are several unique advantages to eigenvector radiosity:

1. The eigenvalues and eigenvectors only need to be calculated once for the conductance matrix $\mathbf{G}$;
2. Changing the initial exitances of the source elements in the environment only affects the initial exitance vector $\mathbf{M}_o$;
3. Changing the surface reflectances only affects the reflectance matrix $\mathbf{R}$ and the vectors $\mathbf{w}_i$; and
4. In many environments, relatively few of the eigenvalues $\lambda_i$ will have significantly large magnitudes.

The fourth advantage has several extremely important consequences. Equations 36 and 37 show that only the largest magnitude eigenvalues will result in significant contributions to $\mathbf{K}^{(r)}$ and so to incremental changes in $\mathbf{M}^{(r)}$. This being the case, only the eigenvectors with the largest magnitude eigenvalues are required to determine the exitance vector $\mathbf{M}$ using Equation 37.

By using a subset of the eigenvectors of $\mathbf{G}$, eigenvector radiosity implicitly uses the approximate conductance matrix $\mathbf{G}'$. Understanding the physical significance of these eigenvectors in terms of dominant paths through the RT network provides a better understanding of how and why eigenvector radiosity works.

DiLaura and Franck[6] used a symmetric form factor matrix representing some 200 random surface elements. They purposely did not model a physical environment in order to avoid having any geometrical properties of the environment skewing their results. What they found was that only a few eigenvalues had absolute magnitudes exceeding 15 percent of the maximum. They also found that $\mathbf{K}^{(r)}$ rapidly approached zero in five or so iterations, and then oscillated about an asymptotic value. (This was probably due to negative values in the equivalent approximate conductance matrix.)

It is important to realize that Equations 36 and 37 converge towards a solution of the approximate radiosity matrix $\mathbf{G}'$ rather than $\mathbf{G}$. Whether this approximation is sufficiently accurate will depend upon the application and how closely $\mathbf{G}'$ approximates $\mathbf{G}$. If more accuracy is required, the solution can be used as the initial exitance vector for further refinement using Gauss-Seidel or Southwell iteration with the radiosity matrix $\mathbf{G}$.

## 9. Jacobi Iteration

We again consider the empty rectangular room example presented in Figure 2, with an initial exitance vector:

$$\mathbf{M}_o = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}^{\mathrm{T}}$$

We can iteratively solve for the final exitance vector $\mathbf{M}$ using Equations 19 or 20, which are both examples of Jacobi iteration[2]. (As shown on p. 380 of Ashdown[1], Jacobi iteration exactly models the physical "bouncing" of light between diffuse reflectance surfaces in an environment.)

Convergence to within a 2-norm relative error of less than 0.1% occurs within 13 iterations. This is typical behavior for Jacobi iteration when applied to the solution of radiative transfer systems.

It is important to recognize that Equations 35 through 37 are simply a reformulation of Jacobi iteration in terms of the eigenpairs of $\mathbf{G}$. Consequently, solving for $\mathbf{M}$ using these equations will produce *exactly* the same convergence behavior if all of the eigenpairs are used.

Using a subset of the eigenpairs to generate the approximate conductance matrix $\mathbf{G}'$ simply means that Jacobi iteration will be applied to a similar RT network. The convergence behavior will be similar, but the final exitance vector $\mathbf{M}$ will be different. For example, substituting $\mathbf{G}'_2$ for $\mathbf{G}$ in Equations 35 through 37 gives:

$$\mathbf{M}'_2 = \begin{bmatrix} 1.2431 & 0.3695 & 0.3695 & 0.3484 & 0.3484 & 0.1322 \end{bmatrix}^{\mathrm{T}}$$

after 12 iterations (again to within a 2-norm relative error of less than 0.1%). This is the same result that was obtained by solving for $\mathbf{M}'_2$ directly using Equation 18.

(As an aside, it should be possible to reformulate Gauss-Seidel and possibly other fast iterative methods in terms of the eigenpairs of $\mathbf{G}$.)

## 10. Eigenanalysis Techniques

If the form factor matrix of a radiative transfer system is known, the eigensystem of its associated conductance matrix can be calculated using standard eigenanalysis techniques. For systems consisting of fewer than 400 or so surface elements, orthogonal similarity transformation techniques such as the QL and QR algorithms[21] are preferred. Once the significant eigenpairs have been determined, the radiosity equations can be trivially solved using Equations 35 through 37.

The situation for large radiative transfer systems is more interesting. Orthogonal similarity transformation techniques quickly become unworkable due to matrix fill-in, and so iterative techniques such as the Lanczos algorithm are necessary.

There are numerous variants of the Lanczos algorithm; two of the more suitable of these are Lanczos with Selective Orthogonalization[22] and Block Lanczos[7,26]. Closely spaced eigenvalues may occur for surface elements that are tightly coupled (that is, where their mutual conductance dominates the RT network), and so Block Lanczos is probably the better choice. (These eigenvalues occur in pairs, and so a block size of two or four is appropriate[21].)

## 11. Southwell Iteration

A more difficult problem arises when the full form factor matrix is not available. This situation occurs for example with progressive radiosity techniques[1,24] that use a variant of Southwell iteration to solve the radiosity equation. Unlike Gauss-Seidel iteration, Southwell iteration requires only one column of the radiosity matrix to be available at

each step. This allows form factors to be calculated “on the fly” and then immediately discarded in order to conserve memory.

(A physical analogy for Southwell iteration is that of surfaces “shooting” radiant flux to all other visible elements in the environment. At each step in the iteration, the surface $k$ with the greatest amount of unsent flux is selected as the shooting surface. This requires determination of the form factors $F_{ki}$, $i = 1,\ldots, n$.)

Southwell iteration often converges to an approximate but useful solution within several hundred to several thousand steps for complex environments with tens of thousands of surface elements. This suggests that a sparse matrix consisting only of those form factors calculated during Southwell iteration can be used to approximate the full form factor matrix. This works well with the Lanczos algorithm, which requires only a user-defined matrix-vector multiply function for sparse matrices.

This approach does not appear to be particularly useful in that the radiosity equation is solved by Southwell iteration, leaving no need for an eigenvector radiosity solution. However, there are many applications where a given radiosity equation must be repeatedly solved for different initial surface exitances and reflectances. After using Southwell iteration to determine an approximate sparse matrix representation of the form factor matrix, we can use eigenvector radiosity to more quickly solve for subsequent initial conditions.

For very large environments with several million elements, several tens of thousands of Southwell iteration steps may be needed to converge to a solution. This may require impractical amounts of memory to store the calculated form factors, even as a sparse matrix. We therefore need to determine which form factors will produce a sparse matrix that acceptably approximates the full matrix.

## 12. Eigenvalue Bounds

A corollary to the Courant-Fischer lemma[2] states that given two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, changes in each eigenvalue of $\mathbf{A}$ for the matrix $\mathbf{A} + \mathbf{B}$ will be between the smallest and largest eigenvalues of $\mathbf{B}$. (See Appendix A for a proof.) Suppose then that we have a sparse conductance matrix $\mathbf{G}_n$ as determined through *n* steps of Southwell iteration. The form factors determined by the next step will produce a matrix $\mathbf{G}_{n+1} = \mathbf{G}_n + \widetilde{\mathbf{G}}$, where $\widetilde{\mathbf{G}}$ is the conductance matrix due the current set of form factors $F_{ki}$.

The bounds of the eigenvalues of $\widetilde{\mathbf{G}}$ are determined by the Gerschgorin Circle theorem[2]. If the largest magnitude eigenvalue of $\widetilde{\mathbf{G}}$ (i.e., its spectral norm) is much less than the spectral norm of $\mathbf{G}_n$, then Equation 16 shows that adding $\widetilde{\mathbf{G}}$ to $\mathbf{G}_n$ will not significantly change this eigensystem, and it can be safely ignored.

Given the structure of $\widetilde{\mathbf{G}}$ (which has only one non-zero row $g_{ik}$ and column $g_{ki}$), the Gerschgorin Circle theorem reduces to:

$$\lambda_{\max} \le A_k \sum_{i=1}^{n} F_{ik} \tag{38}$$

where the bound is the sum of the conductances from surface $k$ to all other surfaces. In physical terms, this bound represents the degree of radiative coupling between surface $k$ and its surrounding environment. For enclosed environments where Equation 6 holds true, the bound is equal to the surface area.

This makes it practical to compute an upper bound on the spectral norm of $\mathbf{G}_{n+1}$ for each step of Southwell iteration. By comparing the eigenvalue bounds to those of

previously calculated sets of form factors, only those sets that will significantly affect the eigensystem of the approximate conductance matrix $\tilde{\mathbf{G}}$ need be cached in memory.

An additional benefit of calculating the eigenvalue bounds for each set of form factors is that it provides an *a priori* estimate of the number of eigenvectors that will need to be calculated using the Lanczos algorithm.

## 13. Locally Dominant Paths

The choice of a "shooting" surface for each step of the Southwell iteration is determined by the maximum amount of unsent radiant flux. Therefore, depending on initial exitance distribution, the sequence of flux transfers between surfaces may not follow the globally dominant paths of the RT network.

A good example is roadway lighting, where the streetlights may directly illuminate only a small portion of the exterior environment. A progressive radiosity solution for this environment may not consider the form factors between surfaces that would be illuminated under (say) daylight conditions.

In terms of the associated RT network, the conductance matrix $\mathbf{G}_n$ will represent locally dominant paths relative to the nodes representing the streetlights. This may not be a concern if changes in the initial exitance distribution are due to changes in the radiant flux from the existing streetlights. However, changes such as adding street lights in previously unlit areas of the environment may involve portions of the RT network that were not considered by the initial progressive radiosity solution. This may require updating $\mathbf{G}_n$ using progressive radiosity.

The situation with interior lighting is different in that interreflections will generally cause the entire environment to be illuminated. In terms of the associated RT network then, the conductance matrix $\mathbf{G}_n$ will usually represent globally dominant paths. Changing the initial exitance distribution or surface reflectances should not in general be influenced by the approximation of $\mathbf{G}_n$ to $\mathbf{G}$ when solving the associated radiosity equation.

## 14. Numerical Experiments

While an empty rectangular room is useful for the purposes of elucidating the principles of eigenvector radiosity, it is obviously not representative of the complex environments that are of interest to architects and lighting designers.

To investigate the potential of eigenvector radiosity as a lighting design tool, a medium-size environment (Figure 4) consisting of 1,192 surface elements was rendered using a progressive radiosity rendering program (Figure 5). The conductance matrix $\mathbf{G}_n$ was assembled during this process by caching 100 sets of form factors with the largest eigenvalue bounds, producing a matrix with 8.3% fill-in. This process took 115 seconds of CPU time on a 450 MHz Pentium desktop computer, with 1.0% convergence reached in 2,228 steps.

The eigensystem of $\mathbf{G}_n$ was then computed using a Block Lanczos eigensolver. Figure 6 shows the distribution of both the eigenvalue bounds and calculated eigenvalues for the first 100 eigenvectors. There were 35 eigenvalues with magnitudes greater than 10% of the maximum eigenvalue (2.9% of the total), and 78 eigenvalues with magnitudes greater than 5% of the maximum (6.5% of the total).

This environment was then rendered a second time using Equations 35 through 37 with 78 eigenvectors (Figure 7). Convergence to 0.1% was reached in 14 iterations, and required only 0.60 seconds of CPU time.

This approximate is surprisingly close to the progressive radiosity solution shown in Figure 5. There are some difference in shading on the screen and walls, but otherwise the results are almost identical.

It is important to remember that the eigenvector radiosity solution can be further refined using progressive radiosity techniques. Because it generates an initial radiant exitance vector $\mathbf{M}_o$ that is close to the final exitance vector $\mathbf{M}$, progressive radiosity should converge very quickly.

## 15. Applications

There are a number of potential applications for eigenvector radiosity. Architectural visualization and lighting design are obvious examples where it is often necessary to determine the effects of changes to surface colors (such as carpeting and painted walls). Once an initial progressive radiosity solution has been calculated, eigenvector radiosity can be used to calculate an approximate exitance distribution for any changes to the environment. This distribution can then be used as the initial exitance distribution for further progressive radiosity calculations to refine the solution.

Another example is daylight analysis. The initial surface exitance distribution for Equation 35 is due to the direct radiant flux contributions from the light sources, which may include direct sunlight and diffuse daylight. Given progressive radiosity solutions for a few selected solar positions and an assumed sky condition, eigenvector radiosity can be used to quickly determine daylight distributions for any intermediate solar position.

Eigenvector radiosity can similarly be applied to thermal engineering analysis, particularly spacecraft and satellite design where radiant heat transfer calculations must be performed for all possible solar orientations.

## 16. Related Work

The eigenvector radiosity approach presented in this paper is an extension of the earlier work of DiLaura and Franck[6]. However, eigenanalysis of radiative transfer systems was first introduced in the context of the radiosity matrix (or more generally, the radiosity kernel for continuous systems) by Moon[13], and later expanded upon by Moon and Spencer[14]. Their interest was in analytic solutions to the radiosity equation.

Koenderink and van Doorn[10] and Langer[12] examined radiosity kernel eigenanalysis from the perspective of computer vision research, in which it was shown that the eigenvectors of the radiosity kernel represent interreflection modes that are invariant with respect to the initial distribution of irradiance.

Baranoski[3], Baranoski et al.[4], and Ramamoorthi[23] attempted to elucidate the physical significance of the radiosity matrix eigenvectors. However, their work did not provide any useful procedural techniques. (Baranoski[3] also noted that a singular value decomposition approach[8] could be applied to obtain a low rank approximation of the radiosity matrix.)

Elements of the proposed eigenvector radiosity approach therefore appeared in the work of DiLaura and Franck[6] and Baranoski[3]. However, the physical significance of the eigensystem can only be understood within the context of radiative transfer networks and their conductance matrices.

## 17. Discussion

It has been shown that eigenanalysis may be used to decompose a radiative transfer network, and that each of its linearly independent components represents a physically realizable radiative transfer network.

It has been further shown that the radiosity equation can be iteratively solved using a subset of the eigendecomposition of its associated conductance matrix, and that

progressive radiosity techniques (in particular, Southwell iteration) can be modified to provide approximate conductance matrices without the need to calculate the full form factor matrix. These are potentially useful results for architectural visualization, lighting design and analysis, thermal engineering, and possibly other fields.

It may be possible to apply eigenvector radiosity to the acoustic domain. There are complicating issues such as sound velocity, media absorption, phase cancellation, and diffraction. However, eigenanalysis of the equivalent resistive circuit network for an acoustic environment may identify the dominant acoustic paths and so simplify the problem of calculating diffuse interreflections.

More generally, linear resistive networks can be used to model neural networks, gas distribution systems, communication networks, transportation systems, nuclear reactors, and structural frameworks (see for example Kron[11]). Because the eigenvector radiosity technique developed in Section 8 relies only on network conductances, it should be applicable to all of these disciplines.

Finally, O'Brien and Bobco[18] and O'Brien and Gomez[19] extended the RT network analogy to include semi-diffuse and specular surfaces, while Oppenheim[20] considered participating media. Future work will investigate whether these extended analogies can be incorporated into eigenvector radiosity.

## 18. Acknowledgments

The results presented in this paper benefited from the careful reading and comments of my thesis advisors Drs. Wolfgang Heidrich and Paul Lalonde, Department of Computer Science, University of British Columbia. I am also indebted to Drs. Chris Jackson and Adam Harris-Harsanyi of Maya Heat Transfer Technologies (Montreal, PQ) for discussions on thermal analysis techniques that led to the development of the conductance network analogy.

Last and most important to me, this paper is dedicated to the memory of my friend and mentor, Dr. Alain Fournier.

## 19. References

1. Ashdown, I. 1994. *Radiosity: A Programmer's Perspective*. New York, NY: John Wiley & Sons.
2. Axelsson, O. 1996. *Iterative Solution Methods*. Cambridge, UK: Cambridge University Press.
3. Baranoski, G. V. G. 1998. *Biologically and Physically-Based Rendering of Natural Scenes*. PhD thesis. Department of Computer Science, University of Calgary, Alberta.
4. Baranoski, G. V. G., R. Bramley, and J. G. Rokne. 1997. "Eigen-analysis for Radiosity Systems," Proceedings of the Sixth International Conference on Computational Graphics and Visualization (Compugraphics '97), pp. 193–201.
5. Deif, A. S. 1991. *Advanced Matrix Theory for Scientists and Engineers, Second Edition*. New York, NY: Abacus Press.
6. DiLaura, D. L., and P. J. Franck. 1993. "On Setting Up and Solving Large Radiative Transfer Systems," Journal of the Illuminating Engineering Society 22(2):3–7 (Summer).
7. Golub, G. H., and R. Underwood. 1977. "The Block Lanczos Method for Computing Eigenvalues," in *Mathematical Software III*, J. Rice, ed. New York, NY: Academic Press.
8. Golub, G. N., and C. F. Van Loan. 1996. *Matrix Computations, Third Edition.* Baltimore, MD: John Hopkins University Press.

9. Howell, J. R. 1982. *A Catalog of Radiation Configuration Factors*. New York, NY: McGraw-Hill.
10. Koenderink, J. J., and A. J. van Doorn. 1983. “Geometrical Modes as a General Method to Treat Diffuse Interreflections in Radiometry,” Journal of the Optical Society of America 73(6):843–850.
11. Kron, G. 1963. *Diakoptics: The Piecewise Solution of Large-Scale Systems*. London, UK: MacDonald.
12. Langer, M. S. 1999. “When Shadows Become Interreflections,” International Journal of Computer Vision 34(2/3):193–204.
13. Moon, P. 1940. “On Interreflections,” Journal of the Optical Society of America 30(5):195–205.
14. Moon, P., and D. E. Spencer. 1981. *The Photic Field*. Cambridge, MA: MIT Press.
15. Neumann, L. 1994. “New Efficient Algorithms with Positive Definite Radiosity Matrix,” Proceedings of the Fifth Eurographics Workshop on Rendering, Darmstadt, Germany, pp. 219–237.
16. Nievergelt, Y. 1997. “Making Any Radiosity Matrix Symmetric Positive Definite,” Journal of the Illuminating Engineering Society 26(1):165–171.
17. O’Brien, P. F. 1955. “Interreflections in Rooms by a Network Method,” Journal of the Optical Society of America 45(6):419–424.
18. O’Brien, P. F., and R. P. Bobco. 1964. “Interreflections in Mirrored Rooms,” Illuminating Engineering 59(5):337–344.
19. O’Brien, P. F., and A. V. Gomez. 1967. “Luminous Transfer in Rooms with Semidiffuse-Specular Surfaces,” Illuminating Engineering 62(4):180–186.
20. Oppenheim, A. K. 1956. “Radiation Analysis by the Network Method,” Transactions of the American Society of Mechanical Engineers 78(1):725–735.
21. Parlett, B. N. 1998. *The Symmetric Eigenvalue Problem*. Philadelphia, PA: SIAM.
22. Parlett, B. N., and D. S. Scott. 1979. “The Lanczos Algorithm with Selective Orthogonalization,” Mathematics of Computation 33(145):217–238.
23. Ramamoorthi, R. 1999. *Eigenmodes of Scene Geometry*. Unpublished manuscript. Stanford Computer Graphics Laboratory. Stanford, CA: Stanford University.
24. Sillion, F. X., and C. Puech. 1994. *Radiosity & Global Illumination*. San Francisco, CA: Morgan Kaufmann.
25. Thomas, R. E., and A. J. Rosa. 1994. *The Analysis and Design of Linear Circuits*. Englewood Cliffs, NJ: Prentice-Hall.
26. Underwood, R. 1975. *An Iterative Block Lanczos Method for the Solution of Large Sparse Symmetric Eigenproblems*. PhD thesis. Technical Report STAN-CS-75-496. Stanford, CA:Stanford University.
27. Yamauti, Z. 1926. “The Light Flux Distribution of a System of Interreflecting Surfaces,” Journal of the Optical Society of America 13(5):561–571.

## Appendix A

**Theorem**: Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices with order *n*, and let $\lambda_i(\mathbf{A})$, $\lambda_i(\mathbf{A}+\mathbf{B})$ denote the *i*th eigenvalue, where the eigenvalues are ordered according to $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$. Then:

$$\lambda_i(\mathbf{A}) + \lambda_{\min}(\mathbf{B}) \le \lambda_i(\mathbf{A}+\mathbf{B}) \le \lambda_i(\mathbf{A}) + \lambda_{\max}(\mathbf{B}) \tag{39}$$

**Proof**: Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ represent the eigenvectors of $\mathbf{A}$. The Courant-Fischer lemma shows that for any **x**, $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$,

$$\lambda_i(\mathbf{A}+\mathbf{B}) \geq \min_{\mathbf{x}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{i-1}} \left(\mathbf{x}^{\mathrm{T}}(\mathbf{A}+\mathbf{B})\mathbf{x}\right) \geq \min_{\mathbf{x}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{i-1}} \left(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\right) + \min\left(\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}\right) \geq \lambda_i(\mathbf{A}) + \lambda_{\min}(\mathbf{B}) \qquad (40)$$

which proves the lower bound, and:

$$\lambda_i(\mathbf{A}+\mathbf{B}) \leq \max_{\mathbf{x}\perp\mathbf{v}_{i+1},\ldots,\mathbf{v}_n} \left(\mathbf{x}^{\mathrm{T}}(\mathbf{A}+\mathbf{B})\mathbf{x}\right) \leq \max_{\mathbf{x}\perp\mathbf{v}_{i+1},\ldots,\mathbf{v}_n} \left(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\right) + \max\left(\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}\right) \leq \lambda_i(\mathbf{A}) + \lambda_{\max}(\mathbf{B}) \qquad (41)$$

which proves the upper bound.

## FIGURE TITLES

Figure 1 – Example radiative transfer network

Figure 2 – Empty rectangular room

Figure 3 – Loop conductance

Figure 4 – Tatami Room (wireframe rendering)

Figure 5 – Tatami Room (progressive radiosity rendering)

Figure 6 – Conductance array eigenvalues

Figure 7 – Tatami Room (eigenvector radiosity rendering)

**Discussion by**

**D.L. DiLaura**
**Department of Civil, Environmental and Architectural Engineering**
**University of Colorado at Boulder**

This work provides another, deeper insight into the significance and computational characteristics of different mathematical formulations of the diffuse radiative transfer problem. The electrical network analogy is helpful, and though it may be no less conceptually abstract than "flux transfer," it affords a physical interpretation of the eigenvectors and eigenvalues characteristic of the matrix that defines the radiative transfer in a diffuse system of surfaces.

The author's insight on this point is very interesting: the radiative transfer in a real system can be considered as comprised of a number of *component* transfer systems; there are as many component systems as there are surfaces in the original system; each component system is described by one eigenvector/value pair from the real system—i.e., each of the matrices governing a component system is defined by one eigenvector and eigenvalue. And, as the author clearly demonstrates, these component systems rank themselves by importance. Do I have it right?

If so, there is a subtle point that might bear emphasizing. These component systems appear *not* to be independent. Evidently, they each "bring flux" to each surface, and the sum of that flux participates in each of the component transfer systems. Is that correct? Or can the process described in the paper's equations 20-26 be arranged so that *independent* components of the final exitance vector are found? Can the component systems be solved independently? Does each component system have a more-or-less simple inverse that gives its solution?

It may be helpful to point out that the conductances in the electrical network analogy are radiative exchange areas, $A_iF_{ij}$, as introduced by Hottel[1]. This might allow the developments in this paper to be considered from a purely flux-transfer point of view, to the benefit of those with an allergy to electrical concepts.

The author has, then, brought us to the brink of the question, "Is there a *flux-transfer interpretation* of these component systems, or rather, the quantities that define their governing matrices?" Each of these matrices can be considered as comprised of *relative* exchange areas (i.e., considering only the eigenvector from which it is made), with each matrix weighted by an eigenvalue. I cannot yet answer this question, and would welcome the author's insight.

An additional small point: I do not understand the statement in the paper that its equation 36, by which an approximation to a system's final exitances are built, is governed by what the author has called the approximate conductance matrix $\mathbf{G'}$. The full rank of eigenvectors/values are being used in equations 20-35, isn't convergence toward the solution of the system defined by $\mathbf{G}$, and not by $\mathbf{G'}$?

## References

1. Hottel, H.C., and A.F. Sarofim, *Radiative Transfer*, McGraw-Hill, 1967, pp. 25-26.

## Response

To answer my mentor's questions in the order presented:

1. Yes, "component transfer systems" is a reasonable physical interpretation. They are closely related to the concept of basis functions, where a complex function can be represented as a weighted sum of a set of basis functions. For example, Fourier

analysis allows a complex signal to be represented as a weighted sum of sines and cosines.

As a matter of interest, it may be possible to express the conductance matrix $\mathbf{G}$ in terms of an arbitrary set of basis functions rather than its eigensystem. The Discrete Cosine Transform (e.g., Jain 1989) is an interesting example in that it minimizes the variance of the transform coefficients of a stationary Markov sequence -- and the form factor matrix describes the probability transition matrix of a Markov process (Feixas et al. 1998). However, calculating the DCT coefficients of a large matrix is considerably more expensive than calculating its eigensystem using a Lanczos eigensolver.

2. Agreed, the component transfer systems are not independent. In general, each system consists of radiative transfer paths between each surface element and all other surface elements visible to it.
3. Two independent component transfer systems would (if I have correctly interpreted the definition of "independent") require that their two sets of surfaces be physically occluded from one another. (An example would be two separate rooms.)

   This however raises an interesting question. Given an arbitrary conductance matrix $\mathbf{G}$ for an *n*-surface environment, Symmetric Schur Decomposition (e.g., Golub and van Loan 1996) states that there exists a real orthogonal $\mathbf{Q}$ such that $\mathbf{Q}^{\mathrm{T}}\mathbf{G}\mathbf{Q} = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1,...,\lambda_n)$. (In physical terms, the mutually orthogonal eigenvectors of $\mathbf{G}$ span an n-dimensional space. It is always possible to find $\mathbf{Q}$ such that the set of eigenvectors is rotated to be aligned with the axes of this space.)

   The question is whether the rotation of $\mathbf{G}$ in *n*-dimensional space has any physical interpretation. If it does, the rotated conductance matrix $\mathbf{\Lambda}$ consists of mutually independent component systems, where each system consists of exactly one surface.
4. Each component system can be solved independently. However, because some of the conductances may be negative, the system may not converge to a solution for some initial conditions (i.e., initial radiative flux distributions). Numerical experiments with random conductance matrices found some examples where the iterative solution diverged.
5. As an antidote for electrical concept allergies, the network can also be modeled as a set of pipes that connect pumps. The analogues to voltage, current, and conductance are pump pressure, water flow, and pipe diameter.
6. Concerning a "flux transfer interpretation," the conductances represent the *ability* of net radiative flux to flow between pairs of surfaces. These are directly related to the environment geometry. The *quantities* of net radiative flux transfer are dependent upon the entire network of conductances and the initial radiative flux distribution. (I suspect however that I do not fully understand the question.)
7. Mea culpa. What I meant to say was that Equations 36 and 37 converge towards the solution of the system defined by $\mathbf{G}'$ if only the eigenvectors comprising the rank-deficient $\mathbf{G}'$ are used in the summation of Equation 36. With the full set of eigenvectors, they of course converge to the solution of $\mathbf{G}$.

As a final aside, Equations 20–37 reformulate Jacobi iteration in terms of the eigensystem of the matrix. There are presumably similar reformulations for other iterative solution methods, such as Gauss-Seidel, conjugate gradient, and MINRES. Whether these would offer any computational advantages for solving radiative transfer systems is currently unknown.

## References


1. Feixas, M., E. del Acebo, and M. Sbert. 1998. "Form Factors and Information Theory," Third International Conference on Computer Graphics and Artificial Intelligence, Limoges, France.
2. Golub, G. H., and C. F. Van Loan. 1996. Matrix Computations, Third Edition. Baltimore, MD: John Hopkins University Press.
3. Jain, A. K. 1989. Fundamentals of Digital Image Processing. Englewood Cliffs, NJ: Prentice Hall.

**Discussion by**

**Jeff Quinlan**
**Lithonia Lighting**

Reading through this paper, it seems like the author is tying two rich topics associated with radiative transfer into one paper. Disassociation of these two topics could have produced two good independent papers. Specifically, the relationship between conductance networks or any numerical approximation of a similar system, and the further development of eigensolutions to the radiative transfer problems.

A critical aspect in solving the system using the eigenvector approximation $\mathbf{G}'$ is the accuracy of the approximation with respect to the radiosity matrix $\mathbf{G}$. What is the best way to determine the accuracy of this matrix? Does the method depend on whether or not the full form factor matrix is known (i.e., is it different for the Southwell approximation)?

Using the same numerical experiment, how do each of the methods that are described in the paper compare? Concerning locally dominant paths, what is the computational expense of the recalculation of the $\mathbf{G}'$ matrix for the same room considering radical changes in the reflectance or the initial illuminance (as might happen in changing a black marble floor to carpet or considering radically different daylighting conditions)?

What is the impact on the computational accuracy of the eigenvalue approximation when using larger surfaces that may encompass a non-uniform initial exitance across the surface face? When using the eigenvector approximation, how long does it take to reduce the error to zero?

**Response**

In my thesis[1], I made the point that the development of eigenvector radiosity presented in this paper required the synthesis of two seemingly unrelated topics: electrical networks and eigenanalysis. They are tied together in this paper not by choice, but by necessity.

For the purpose of radiative transfer analysis, the accuracy of $\mathbf{G}'$ with respect to $\mathbf{G}$ is dependent upon the initial exitance vector $\mathbf{M}_o$. Given the radiative transfer equations $\mathbf{M}_0 = (\mathbf{I} - \mathbf{SG})\mathbf{M}$ and $\mathbf{M}_0 = (\mathbf{I} - \mathbf{SG}')\mathbf{M}'$, the difference $\mathbf{M} - \mathbf{M}'$ represents the per-surface error. The global accuracy could be expressed as (for example) $\|\mathbf{M} - \mathbf{M}'\| / \|\mathbf{M}\|$, but this is a measure of average exitance. We are more often interested in the maximum normalized per-surface error, which is $\max(|m_i - m_i'|)_{i=1}^{n} / \|\mathbf{M}\|$. In any event, the error can only be determined *a posteriori*.

In general however, the Courant-Fischer lemma provides a reasonable indication of the global accuracy of $\mathbf{G}'$ with respect to $\mathbf{G}$, irrespective of the initial exitance vector $\mathbf{M}_o$.

Regarding the choice of methods, it must be noted that Southwell iteration is not an approximation. Like Gauss-Seidel and Jacobi iteration, it will (ignoring roundoff errors) converge to an exact solution. It produces an approximation $\mathbf{M}'_j$ with a monotonically decreasing residual $\mathbf{M} - \mathbf{M}'_j$ for each iterative step *j*.

Constructing a low-rank approximation of the conductance matrix $\mathbf{G}$ using progressive radiosity as described in the paper adds a further layer of approximation to

the eigenvector approximation $\mathbf{G}'$. Again, it is not possible to determine the accuracy of the final exitance vector *a priori*. However, the full form factor matrix will produce a more accurate solution.

Regarding the numerical experiment, no attempt was made to solve for the Tatami Room environment using the full form factor matrix. Storing the matrix would have required 5.4 MB of RAM. The software was designed for much larger environments (typically 50,000 elements), which would require on the order of 10 GB of RAM.

As for locally dominant paths, it would not be necessary to recalculate $\mathbf{G}'$ (which is strictly a function of the environment geometry) for the Tatami Room environment. This is necessary only for large and complex environments that are not fully illuminated.

Finally, the radiative transfer equation is predicated on the assumption of uniform initial exitance across each surface element. While it is possible to represent non-uniform surface exitance distributions (as was done for example in Mr. Quinlan's thesis[2]), this issue has yet to be addressed in the context of eigenvector radiosity.

**References**


1. Ashdown, I. 2001. Eigenvector Radiosity. MSc thesis, Department of Computer Science, University of British Columbia.
2. Quinlan, J. 1994. Calculating Illuminance from Non-Diffuse Sources. MSc thesis, Department of Civil, Environmental, and Architectural Engineering, University of Colorado at Boulder.

**Discussion by**

**Dr. Wolfgang Heidrich**
**Department of Computer Science**
**University of British Columbia**

In Section 6, "Approximate Conductance Networks," the diagonal elements of the conductance matrix are interpreted as physical conductances. Because they are loop conductances, it is assumed that they have no net flow of radiative flux and hence no effect on the radiative transfer equation solution.

This implies that the diagonal elements of the approximate conductance matrix are unconstrained, which seems unlikely. Has the author investigated this issue?

**Response**

My thesis advisor raised this issue at the time of my thesis presentation, and I was unable to answer it then. I have since discovered that my physical interpretation of the network nodes is in need of revision.

Each node *i* represents a physical surface that diffusely reflects some of its incident flux back into the environment. Where the surface is concave, its form factor $F_{ii}$ is non-zero because the surface can "see" itself.

The surface has reflectance $\rho$, which represents the portion of radiant flux that is reflected for a differential area of the surface. However, this does not represent the effective reflectance of the entire surface element.

The radiant flux balance for each reflection is:

| | |
|---|---|
| $\rho(1-F)$ | reflected away from surface |
| $1-\rho$ | absorbed by surface |
| $\rho F$ | interreflected to surface |

which means that the effective reflectance due to the infinite series of interreflections between the surface and itself is:

$$\rho_{eff} = \rho(1-F)\left(1+(\rho F)+(\rho F)^2+(\rho F)^3+\ldots\right)$$

Using the geometric series expansion:

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} r^n, \quad -1 < r < 1$$

we have:

$$\rho_{eff} = \rho(1-F)/(1-\rho F)$$

As the form factor $F_{ii}$ increases from zero to unity, the effective surface reflectance $\rho_{eff}$ tends to zero, as we would expect.

In terms of network theory, this behavior can be represented by modeling a concave surface as an amplifier with positive feedback, and whose gain without feedback is $\rho(1-F)$. The loop conductance then becomes the feedback path between the output and input ports.

Alternatively, recalling that $F_{ii} = g_{ii} / A_{ii}$ each surface reflectance $\rho$ can be replaced with the corresponding effective surface reflectance $\rho_{eff}$, in which case the diagonal elements of the approximate conductance matrix become zero.

With this, it can be seen that the value of the loop conductance determines the effective surface reflectance and so affects the radiative transfer equation solution.

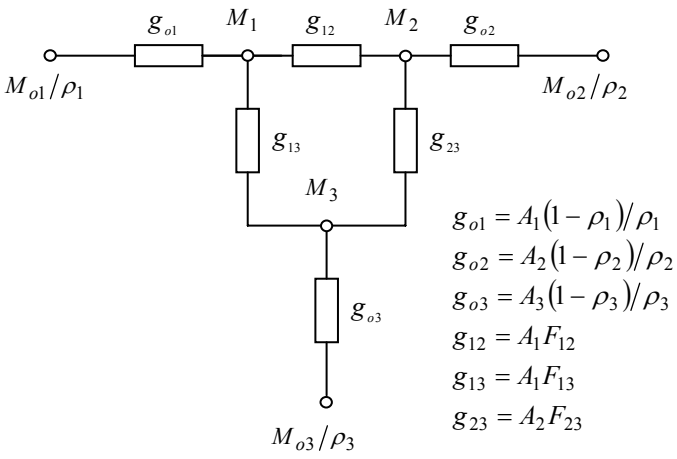


Figure 1

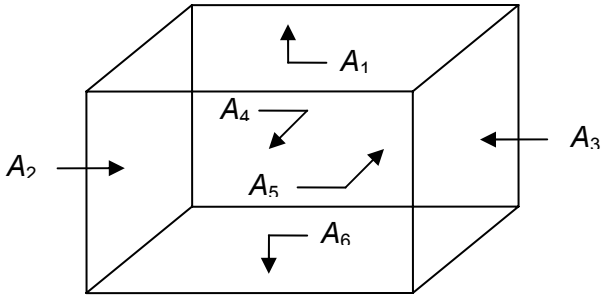


Figure 2

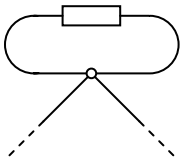

Figure 3

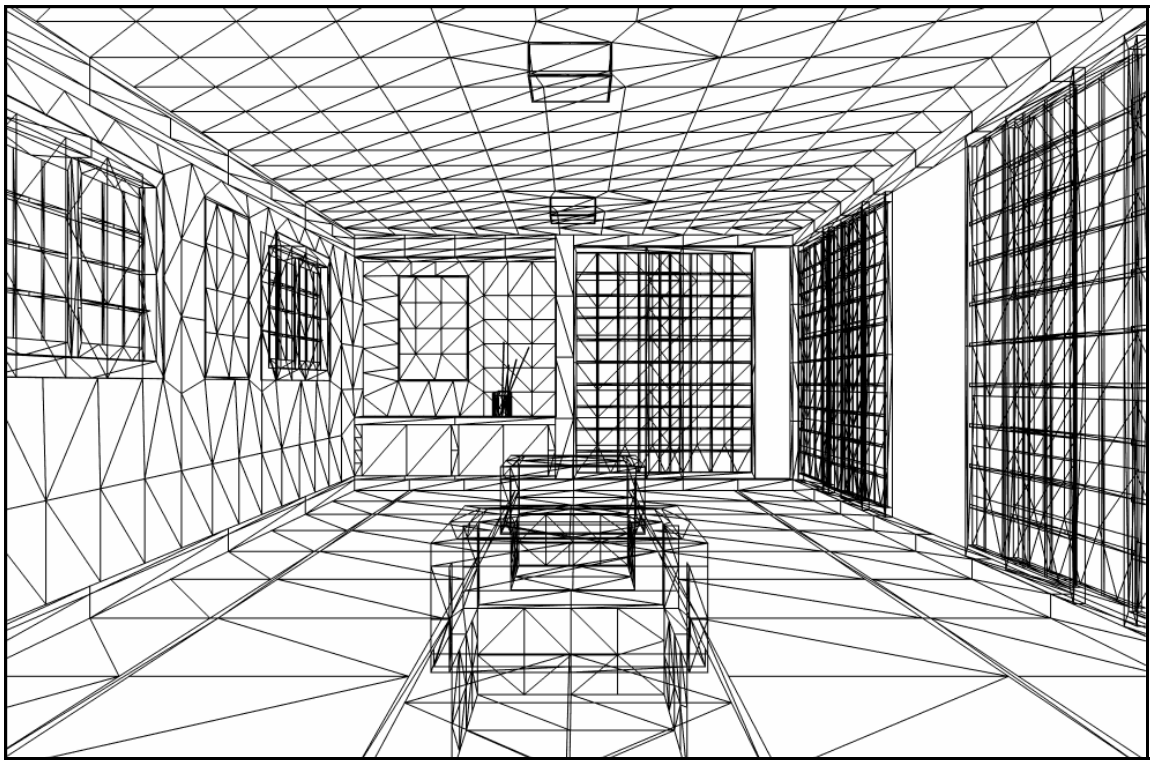

Figure 4

Figure 5

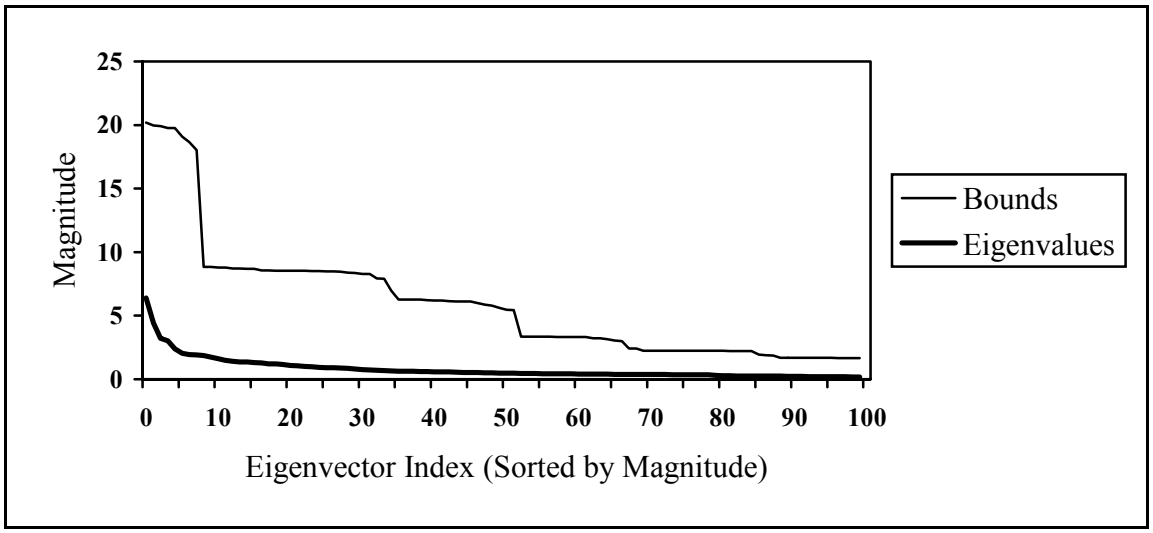


Figure 6

Figure 7