Interner Bericht

Variance reduction for Monte Carlo methods by means of deterministic numerical computation

Stefan Heinrich

275/95

Fachbereich Informatik

Universität Kaiserslautern · Postfach 3049 · D-67653 Kaiserslautern

Variance reduction for Monte Carlo methods by means of deterministic numerical computation

Stefan Heinrich

275/95

Universität Kaiserslautern
AG Numerische Algorithmen
Postfach 30 49
67653 Kaiserslautern

September 1995

Herausgeber: AG Numerische Algorithmen
Leiter: Professor Dr. S. Heinrich

Variance reduction for Monte Carlo methods by means of deterministic numerical computation

Stefan Heinrich
Universität Kaiserslautern
Fachbereich Informatik
Postfach 30 49
67653 Kaiserslautern
email: heinrich@informatik.uni-kl.de

Abstract

A new variance reduction technique for the Monte Carlo solution of integral equations is introduced. It is based on separation of the main part. A neighboring equation with exactly known solution is constructed by the help of a deterministic Galerkin scheme. The variance of the method is analyzed, and an application to the radiosity equation of computer graphics, together with numerical test results is given.

1 Introduction

In [HM93], [Hei94] the complexity of Monte Carlo solution of integral equations was studied. To prove upper bounds, a new variance reduction technique was developed which combined deterministic and stochastic computations in an optimal way. The analysis of [HM93], [Hei94] aimed at sharp estimates of complexity rates, so it was carried out for a model problem - smooth kernels and right-hand sides over the unit cube. In the present paper we develop this method for general domains and non-smooth kernels in such a way that it can be used for Monte Carlo simulations of real processes. We shall present an application to global illumination in computer graphics.
Let us briefly describe the underlying idea. Suppose we have a second integral equation (i. e. second kernel and right-hand side) which is close to our original one. Knowing the solution of the latter, it is possible to improve the accuracy of Monte Carlo estimates of the former considerably. This is the well-known method of control variates or separation of main part. It is described in section 2. Usually the second equation stems from a neighboring problem (maybe simpler one, maybe one already computed). But what to do in the general case, when no such parallel problem suggests itself. That was the idea in [HM93]: Let us first calculate some deterministic approximation to the solution. In high-dimensional problems, this will be only a rough one, as a rule, less precise than Monte Carlo. Now we construct a neighboring problem. We arrange it in such a way

that the approximate solution of our original equation becomes the exact solution of the approximate equation. If that approximate problem is close to the original one, if it can be solved by the help of the same Markov process, and if its involved functions are not too complex, we can use both equations to obtain a considerably improved Monte Carlo solution.

Of course, for concrete simulation problems, this approach has many "if"s. We try to resolve some of them by providing a general analysis of how the proximity of kernels and right-hand sides affects the reduction of variance. This is the contents of section 3. In section 4 we show how the data of a piecewise constant Galerkin approximation can be used to set up a neighboring equation. On the basis of accuracy estimates from section 3 we analyze which parameters of the deterministic approximation influence the reduction of variance. In section 5 we outline how the procedure can be applied to the radiosity equation in computer graphics. Finally, we provide results of numerical experiments based on a simple test equation from two-dimensional ("flatland") radiosity, whose solutions are known exactly.

For general information on Monte Carlo methods, variance reduction, and control variates we refer to [SG69], [Erm71], [Sob73], [EM82], [KW86], [ENS89], [Mik91a], [Mik91b], [Sab91]. Several previously developed variance reduction techniques for integral equations are related to ours in the sense that they are based on control variates, but they lead to entirely different algorithms: [Erm71],ch. 6.2.5, [Spa79], [ES85], [Mik91b],§5.10, [Sab91], ch. 2.2.3, [LW94]. For complexity theory of Monte Carlo methods we refer to [TWW88], [Nov88], [Hei94], [Mat]. The computer graphics background is covered by [CW93], while notions from functional analysis can be found in [DS58].

2 Separation of the main part

Let X be a non-empty set, Σ a σ-algebra of subsets, μ a positive, σ-additive finite measure on (X, Σ). Let $1 \leq s \leq \infty$ and let $L_s(X) = L_s(X, \Sigma, \mu)$ denote the usual Banach space of Σ-measurable s-integrable functions (resp. essentially bounded functions, for $s = \infty$). Consider the integral equation

$$u(x) \;=\; \int_X K(x,y)u(y)\, d\mu(y) + F(x) \tag{1}$$

where F and K are given such that $F \in L_\infty(X)$, K belongs to the Banach space - denoted here by $L_\infty(L_1)$ - of $\Sigma \times \Sigma$ measurable functions on X^2 satisfying

$$\|K\|_{L_\infty(L_1)} := \operatorname*{ess\,sup}_{x \in X} \int_X |K(x,y)|\, d\mu(y) < \infty,$$

and $u \in L_\infty(X)$ is the unknown solution.

Under these conditions the integral operator T_K defined for $f \in L_\infty(X)$ by

$$(T_K f)(x) = \int_X K(x,y) f(y)\, d\mu(y)$$

is a bounded linear operator in $L_\infty(X)$ and its operator norm satisfies

$$\|T_K : L_\infty(X) \to L_\infty(X)\| = \|K\|_{L_\infty(L_1)}.$$

We assume that we are given a function $\Phi \in L_1(X, \Sigma, \mu)$. Instead of the full solution $u(x)$ we seek only the value of the functional Φ at u, that is, the scalar product

$$(u, \Phi) = \int_X u(x)\Phi(x)\, d\mu(x). \tag{2}$$

A standard Monte Carlo procedure to compute (2) is the von Neumann Ulam scheme, which consists in the following. Let $p_0(x)$ and $p(x,y)$ be the densities of initial distribution and transition probability, respectively, of an absorbing Markov chain on X. We assume that $p_0(x)$ and $p(x,y)$ are measurable on X and $X \times X$, respectively, $p_0(x) \geq 0$, $p(x,y) \geq 0$ $(x, y \in X)$,

$$\int_X p_0(x)\, d\mu(x) = 1 \tag{3}$$

and

$$\int_X p(x,y)\, d\mu(y) = 1 - q(x) \tag{4}$$

with $q(x) \geq 0$ being the probability of absorption at the point x. We suppose that the spectral radius of T_p in $L_\infty(X)$ is less than 1. This guarantees that almost all realizations of the Markov chain are of finite length. To proceed with the von Neumann Ulam scheme we suppose

$$\begin{aligned} \mu\{\Phi(x) \neq 0 \quad \text{and} \quad p_0(x) = 0\} &= 0 \\ \mu \times \mu\{K(x,y) \neq 0 \quad \text{and} \quad p(x,y) = 0\} &= 0 \\ \mu\{F(x) \neq 0 \quad \text{and} \quad q(x) = 0\} &= 0. \end{aligned}$$

For our purposes it is convenient to represent

$$\begin{aligned} \Phi(x) &= \varphi(x) p_0(x) \\ K(x,y) &= k(x,y) p(x,y) \\ F(x) &= f(x) q(x). \end{aligned}$$

So our original equation (1) becomes

$$u = T_{kp} u + fq. \tag{5}$$

Given a realization

$$\xi = (x_0, x_1, \ldots, x_m) \tag{6}$$

of the random walk, we define a random variable η by

$$\eta(k, f, \xi) = \varphi(x_0)k(x_0, x_1)\dots k(x_{m-1}, x_m)f(x_m). \tag{7}$$

If the spectral radius of $T_{|K|} = T_{|k|p}$ in $L_\infty(X)$ is less than 1, then μ has finite first moment and

$$(u, \Phi) = \mathbb{E}\eta(k, f, \xi) \tag{8}$$

and the Monte Carlo approximation to (u, Φ) is given by

$$\frac{1}{N}\sum_{i=1}^{N}\eta(k, f, \xi_i), \tag{9}$$

where the $(\xi_i)_{i=1}^N$ are independent realizations of the Markov chain. This accomplishes the von Neumann Ulam scheme. Now let us assume that we have approximations v to u, h to k and g to f such that the corresponding integral equation

$$v = T_{hp}v + gq$$

holds exactly. Let us put

$$\zeta(k, f, h, g, \xi) = (v, \Phi) + \eta(k, f, \xi) - \eta(h, g, \xi). \tag{10}$$

Then (u, Φ) is approximated by

$$\frac{1}{N}\sum_{i=1}^{N}\zeta(k, f, h, g, \xi_i) = (v, \Phi) + \frac{1}{N}\sum_{i=1}^{n}(\eta(k, f, \xi_i) - \eta(h, g, \xi_i)). \tag{11}$$

This is known as the method of control variates or separation of main part for integral equations (see, e. g., [SG69], ch. 3.9, [Erm71], ch. 6, §2.5). The mean square error of approximation of (u, Φ) by (11) is given by

$$\mathbb{E}((u, \Phi) - \frac{1}{N}\sum_{i=1}^{N}\zeta(k, f, h, g, \xi_i))^2 = \frac{\text{Var }(\zeta)}{N}.$$

It is the main goal of section 3 to give a detailed analysis of the variance of ζ, its dependence on the kernels and right hand sides involved, and in particular, the rate of decrease of Var (ζ) as h approximates k and g approximates f. We shall carry out this analysis for a fixed Markov chain, satisfying the assumptions stated above. We shall also fix $\Phi = \varphi p_0 \in L_1(X)$ and assume, in addition,

$$\varphi^2 p_0 \in L_1(X).$$

Variable parameters will be the functions k, h, f, g, and it is convenient to introduce the following classes: Fix $\alpha > 0$, $0 < \gamma < 1$, $n_0 \in \mathbb{N}$ and let $\mathcal{K} = \mathcal{K}(\alpha, \gamma, n_0)$ be the set of all $k \in L_\infty(X^2)$ satisfying

$$\|k\|_{L_\infty(X^2)} \leq \alpha \tag{12}$$

and

$$\|T_{k^2p}^{n_0} : L_\infty(X) \to L_\infty(X)\| \le \gamma. \tag{13}$$

Given $\theta > 0$ we let $\mathcal{F} = \mathcal{F}(\theta)$ be the class of all measurable on X functions f with $f^2 q \in L_\infty(X)$ and

$$\|f^2 q\|_{L_\infty(X)} \le \theta. \tag{14}$$

Observe that this and (4) imply

$$\|fq\|_{L_\infty(X)} \le \|fq^{1/2}\|_{L_\infty(X)} \le \theta^{1/2}.$$

It is well-known that for $k \in \mathcal{K}$ and $f \in \mathcal{F}$, the random variable η is well-defined and possesses finite second moment (see also Lemma 1 below). Hence, if $k, h \in \mathcal{K}$ and $f, g \in \mathcal{F}$, then the variable ζ is also well-defined and of finite variance. For the further estimates we need the following results about $\mathcal{K}$. These facts are elementary and mostly known. We list them, together with an outline of proof, for the sake of completeness.

Lemma 1 *Let $\alpha > 0$, $0 < \gamma < 1$, $n_0 \in \mathbb{N}$. If $k \in \mathcal{K}(\alpha, \gamma, n_0)$, then*

(i) $I - T_{k^2p}$ *is invertible in* $L_\infty(X)$ *and* $\|(I - T_{k^2p})^{-1} : L_\infty(X) \to L_\infty(X)\| \le \beta_0$, *where* $\beta_0 = (1-\gamma)^{-1} \sum_{j=0}^{n_0-1} \alpha^{2j}$,

(ii) $\|kp\|_{L_\infty(L_1)} \le \|k^2 p\|_{L_\infty(L_1)}^{1/2} \le \alpha$,

(iii) $\|T_{kp}^{n_0} : L_\infty(X) \to L_\infty(X)\| \le \gamma$,

(iv) $I - T_{kp}$ *is invertible in* $L_\infty(X)$ *and* $\|(I - T_{kp})^{-1} : L_\infty(X) \to L_\infty(X)\| \le \beta_1$, *where* $\beta_1 = (1-\gamma)^{-1} \sum_{j=0}^{n_0-1} \alpha^j$.

If $k, h \in \mathcal{K}(\alpha, \gamma, n_0)$, then

(v) $\|T_{khp}^{n_0} : L_\infty(X) \to L_\infty(X)\| \le \gamma$, *and* $\|(I - T_{khp})^{-1} : L_\infty(X) \to L_\infty(X)\| \le \beta_0$.

Proof: (i) is just the Neumann series

$$(I - T_{k^2p})^{-1} = (I - T_{k^2p}^{n_0})^{-1} \sum_{j=0}^{n_0-1} T_{k^2p}^j,$$

observing also that

$$\|k^2 p\|_{L_\infty(L_1)} \le \alpha^2$$

because of (4). (ii) is Hölder's inequalitiy:

$$\int_X |k(x,y)|p(x,y)\ d\mu(y) \le \left(\int_X k(x,y)^2 p(x,y)\ d\mu(y)\right)^{1/2} \left(\int_X p(x,y) d\mu(y)\right)^{1/2}.$$

To see (iii), note that

$$\begin{aligned} &\|T_{kp}^{n_0} : L_\infty(X) \to L_\infty(X)\| \\ \le\ &\operatorname*{ess\,sup}_{x_0 \in X} \int_{X^{n_0}} \prod_{i=0}^{n_0-1} (|k(x_i, x_{i+1})|p(x_i, x_{i+1}))\ d\mu(x_1) \dots d\mu(x_{n_0}). \end{aligned}$$

which can be estimated by Hölder's inequality in a similar way as above. (iv) follows from (ii) and (iii) as in (i). Finally, (v) is a consequence of

$$2 \prod_{i=0}^{n_0-1} |k(x_i, x_{i+1})h(x_i, x_{i+1})| \le \left(\prod_{i=0}^{n_0-1} k(x_i, x_{i+1})\right)^2 + \left(\prod_{i=0}^{n_0-1} h(x_i, x_{i+1})\right)^2.$$

This proves the Lemma.

Remarks:

1. For the Lemma to hold true, it suffices to assume

$$\|k^2 p\|_{L_\infty(L_1)} \le \alpha^2 \tag{15}$$

instead of the stronger condition (12).

2. Although we restricted ourselves to $\Phi \in L_1(X)$, our analysis also includes e. g. delta functions. Suppose X is endowed with a metric and we consider continuous (or piecewise continuous) functions on X instead of general L_∞ functions. Then point-values are correctly defined and constitute continuous linear functionals. In that case we have

$$(u, \delta_{x_0}) = u(x_0) = F(x_0) + \int_X K(x_0, y)u(y)\ d\mu(y),$$

and the problem reduces to the computation of (u, Φ), where $\Phi = K(x_0, y)$, and it is assumed that $K(x_0, \cdot) \in L_1(X)$.
Setting

$$p_0(y) = p(x_0, y)/(1 - q(x_0)).$$

(provided $q(x_0) \neq 1$), we essentially arrive at the standard process of computing $u(x_0)$ which simply starts in x_0 with probability 1, then moves to the next state with density $p(x_0, \cdot)$ etc.

3 Variance analysis

Following the arguments of variance computing for the standard von Neumann Ulam scheme, ([Erm71] ch. 6, §2) we first give an exact expression for the variance of ζ.

Proposition 2 *Let $\alpha, \theta > 0$, $0 < \gamma < 1$, $n_0 \in \mathbb{N}$ and assume that $k, h \in \mathcal{K}(\alpha, \gamma, n_0)$, $f, g \in \mathcal{F}(\theta)$. Then the variance of $\zeta(k, f, h, g, \xi)$ is given by*

$$\begin{aligned} \mathrm{Var}\ (\zeta) &= ((I - T_{k^2p})^{-1}(f^2q) - 2(I - T_{khp})^{-1}(fgq) + (I - T_{h^2p})^{-1}(g^2q), \varphi^2 p_0) \\ &\quad -((I - T_{kp})^{-1}(fq) - (I - T_{hp})^{-1}(gq), \varphi p_0)^2. \end{aligned} \tag{16}$$

Proof: We have

$$\mathrm{Var}(\zeta) = \mathrm{Var}(\eta(k, f, \xi) - \eta(h, g, \xi)), \tag{17}$$

and

$$\mathbb{E}(\eta(k, f, \xi) - \eta(h, g, \xi))^2 = \sum_{m=0}^{\infty} \int_{X^{m+1}} \Lambda_m(x_0, \ldots, x_m)\ d\Theta(x_0, \ldots, x_m),$$

where

$$\Lambda_m(x_0, \ldots, x_m) = \varphi(x_0)^2 (f(x_m) \prod_{i=0}^{m-1} k(x_i, x_{i+1}) - g(x_m) \prod_{i=0}^{m-1} h(x_i, x_{i+1}))^2 \tag{18}$$

and

$$d\Theta(x_0, \ldots, x_m) = p_0(x_0) q(x_m) \prod_{i=0}^{m-1} p(x_i, x_{i+1})\ d\mu(x_0) \ldots d\mu(x_m)$$

(in case $m = 0$ we replace the product $\prod_{i=0}^{m-1}$ by the factor 1). Consequently,

$$\begin{aligned} \Lambda_m(x_0, \ldots, x_m) &= \varphi(x_0)^2 (f(x_m)^2 \prod_{i=0}^{m-1} k(x_i, x_{i+1})^2 \\ &\quad -2 f(x_m) g(x_m) \prod_{i=0}^{m-1} (k(x_i, x_{i+1}) h(x_i, x_{i+1})) + g(x_m)^2 \prod_{i=0}^{m-1} h(x_i, x_{i+1})^2) \end{aligned}$$

from which we get that

$$\begin{aligned} &\mathbb{E}(\eta(k, f, \xi) - \eta(h, g, \xi))^2 \\ &= \sum_{m=0}^{\infty} (T_{k^2p}^m (f^2 q) - 2 T_{khp}^m (fgq) + T_{h^2p}^m (g^2 q), \varphi^2 p_0) \\ &= ((I - T_{k^2p})^{-1}(f^2 q) - 2(I - T_{khp})^{-1}(fgq) + (I - T_{h^2p})^{-1}(g^2 q), \varphi^2 p_0). \end{aligned} \tag{19}$$

The absolute convergence of the series involved here follows from the assumptions and Lemma 1. Combining this with (8) and (17) yields the result.

Remark: Proposition 2 holds true if we replace (12) by (15). Only in the subsequent results we make use of the stronger requirement (12).

Although formula (16) is exact, it is not convenient for us since it does not show explicitly the dependence of the variance on the proximity of h to k and g to f. We shall therefore provide further estimates, which will fulfill this requirement.
Let us first introduce some notation We put

$$\|T_{kp} - T_{hp} : L_\infty(X) \to L_\infty(X)\| = \delta(k,h) \tag{20}$$

$$\|T_{(k-h)^2 p} : L_\infty(X) \to L_\infty(X)\| = \varepsilon(k,h)^2 \tag{21}$$

$$\|(f-g)^2 q\|_{L_\infty(X)} = \rho(f,g)^2. \tag{22}$$

Observe that

$$\begin{aligned} \delta(k,h) &= \operatorname*{ess\,sup}_{x\in X} \int_X |k(x,y) - h(x,y)| p(x,y)\, d\mu(y) \\ &\leq \operatorname*{ess\,sup}_{x\in X} \left(\int_X (k(x,y) - h(x,y))^2 p(x,y)\, d\mu(y) \right)^{1/2} = \varepsilon(k,h). \end{aligned}$$

The following theorem shows that the variance reduction is at least proportional to the square of the approximation errors of k by h and f by g, in the respective norms.

Theorem 3 *Let $\alpha, \theta > 0$, $0 < \gamma < 1$ and $n_0 \in \mathbb{N}$. Then there exists a constant $c > 0$ such that for all $k, h \in \mathcal{K}(\alpha, \gamma, n_0)$ and $f, g \in \mathcal{F}(\theta)$,*

$$\mathrm{Var}(\zeta(k,f,h,g,\xi)) \leq c(\varepsilon(k,h)^2 + \rho(f,g)^2).$$

Proof: From (17) we get

$$\begin{aligned} \mathrm{Var}(\zeta) &\leq \mathbb{E}(\eta(k,f,\xi) - \eta(h,g,\xi))^2 \\ &\leq 2\,\mathbb{E}(\eta(k,f,\xi) - \eta(h,f,\xi))^2 + 2\,\mathbb{E}(\eta(h,f,\xi) - \eta(h,g,\xi))^2 \\ &= 2(w_1 + w_2, \varphi^2 p_0) \end{aligned} \tag{23}$$

where, according to (19), $w_1, w_2 \in L_\infty(X)$ are given by

$$w_1 = ((I - T_{k^2 p})^{-1} - 2(I - T_{khp})^{-1} + (I - T_{h^2 p})^{-1})(f^2 q) \tag{24}$$

$$w_2 = (I - T_{h^2 p})^{-1}((f-g)^2 q). \tag{25}$$

To abbreviate formulas, we put

$$\begin{aligned} A &= (I - T_{k^2p}) \\ B &= (I - T_{khp}) \\ C &= (I - T_{h^2p}). \end{aligned}$$

Then we have

$$\begin{aligned} & A^{-1} - 2B^{-1} + C^{-1} \\ = \; & A^{-1}(B - A)B^{-1} + C^{-1}(B - C)B^{-1} \\ = \; & A^{-1}(B - A)B^{-1} + A^{-1}(B - C)B^{-1} + (C^{-1} - A^{-1})(B - C)B^{-1} \\ = \; & A^{-1}(B - A)B^{-1} + A^{-1}(B - C)B^{-1} + A^{-1}(A - C)C^{-1}(B - C)B^{-1} \\ = \; & A^{-1}[2B - A - C + (A - C)C^{-1}(B - C)]B^{-1}. \end{aligned}$$

Hence

$$w_1 = (I - T_{k^2p})^{-1}(T_{(k-h)^2p} + T_{(h^2-k^2)p}(I - T_{h^2p})^{-1}T_{(h-k)hp})(I - T_{khp})^{-1}(f^2q). \quad (26)$$

In view of (12) we have

$$\|T_{(h-k)hp} : L_\infty(X) \to L_\infty(X)\| = \|(h - k)hp\|_{L_\infty(L_1)} \leq \alpha\delta(k, h) \quad (27)$$

and similarly

$$\|T_{(h^2-k^2)p} : L_\infty(X) \to L_\infty(X)\| = \|(h + k)(h - k)p\|_{L_\infty(L_1)} \leq 2\alpha\delta(k, h). \quad (28)$$

It follows from Lemma 1, (21), (27) and (28) that

$$\begin{aligned} \|w_1\|_{L_\infty(X)} &\leq \beta_0(\varepsilon(k, h)^2 + 2\alpha\delta(k, h)\beta_0\alpha\delta(k, h))\beta_0\theta \\ &\leq \beta_0^2\theta(1 + 2\alpha^2\beta_0)\varepsilon(k, h)^2. \end{aligned} \quad (29)$$

and, recalling (22) and (25),

$$\|w_2\|_{L_\infty(X)} \leq \beta_0\rho(f, g)^2. \quad (30)$$

By (23),

$$\mathrm{Var}(\zeta) \leq 2(\|w_1\|_{L_\infty(X)} + \|w_2\|_{L_\infty(X)})\|\varphi^2 p_0\|_{L_1(X)}$$

which together with (29) and (30) proves the theorem.

Remarks: 1. The constant c depends only on $\alpha, \gamma, n_0, \theta$ and on $\|\varphi^2 p_0\|_{L_1(X)}$.

2. We do not have to demand $h \in \mathcal{K}$. A standard perturbation argument yields the following: Given $\alpha > 0$, $0 < \gamma < \gamma' < 1$ and n_0, there exists a $\delta_0 > 0$ such that whenever $k \in \mathcal{K}(\alpha, \gamma, n_0)$ and $h \in L_\infty(X^2)$ is such that $\|h\|_{L_\infty(X^2)} \leq \alpha$ and $\delta(k, h) \leq \delta_0$, then $h \in \mathcal{K}(\alpha, \gamma', n_0)$.

Corollary 4 *Given $\alpha, \theta > 0$, $0 < \gamma < 1$, $n_0 \in \mathbb{N}$, there exist constants $\delta_0 > 0$ and $c > 0$ such that for all $k \in \mathcal{K}(\alpha, \gamma, n_0)$ and $h \in L_\infty(X^2)$ with $\|k - h\|_{L_\infty(X^2)} \leq \delta_0$, and for all $f, g \in L_\infty(X)$ with $\|f\|_{L_\infty(X)}$, $\|g\|_{L_\infty(X)} \leq \theta$, the Monte Carlo scheme (7) is well-defined and*

$$\mathrm{Var}(\zeta) \leq c(\|k - h\|^2_{L_\infty(X^2)} + \|f - g\|^2_{L_\infty(X)}).$$

This follows readily from Theorem 3 and the remark above. Corollary 4 extends the upper estimate in [HM93], Prop. 1, where the case $X = [0,1]^d$, $p(x,y) \equiv 1 - q_0$ was considered.

Sometimes it might be hard to get an estimate of $\delta(k,h)$, $\varepsilon(k,h)$ or $\|k - h\|_{L_\infty(X^2)}$ because of the supremum involved in their definitions. This occurs, for example, when the radiosity equation (see below) is considered in domains with corners. By strengthening the assumptions on φ, k and h, we can pass to the L_2-norm. We shall assume that $\varphi^2 p_0 \in L_\infty(X)$, and

$$c_p = \operatorname*{ess\,sup}_{y \in X} \int_X p(x,y)\, d\mu(x) \quad < \infty. \tag{31}$$

Hence if $k \in L_\infty(X^2)$, then $T_{k^2 p}$ is bounded also as an operator in $L_1(X)$. Let us introduce the following class. Given $\alpha > 0$, $\beta > 1$, $0 < \gamma < 1$ and $n_0 \in \mathbb{N}$, define

$$\mathcal{K}'(\alpha, \beta, \gamma, n_0) = \{k \in \mathcal{K}(\alpha, \gamma, n_0) : \quad I - T_{k^2 p} \text{ is invertible in } L_1(X) \text{ and } \|(I - T_{k^2 p})^{-1} : L_1(X) \to L_1(X)\| \leq \beta\}.$$

By Lemma 1, each $k \in \mathcal{K}(\alpha, \gamma, n_0)$ satisfies

$$\|(I - T_{k^2 p})^{-1} : L_\infty(X) \to L_\infty(X)\| \leq \beta_0,$$

hence the Riesz-Thorin interpolation theorem (see [Tri78]) implies that for $k \in \mathcal{K}'(\alpha, \beta, \gamma, n_0)$

$$\|(I - T_{k^2 p})^{-1} : L_2(X) \to L_2(X)\| \leq (\beta \beta_0)^{1/2}. \tag{32}$$

Let us finally denote

$$\begin{aligned} \sigma(k,h)^2 &= \int_X \int_X (k(x,y) - h(x,y))^2 p(x,y)\, d\mu(x)\, d\mu(y) \\ \tau(f,g)^2 &= \int_X (f(x) - g(x))^2 q(x)\, d\mu(x). \end{aligned}$$

Theorem 5 *Given $\alpha, \theta > 0$, $\beta > 1$, $0 < \gamma < 1$, $n_0 \in \mathbb{N}$, there exists a constant $c > 0$ such that for $k, h \in \mathcal{K}'(\alpha, \beta, \gamma, n_0)$ and $f, g \in \mathcal{F}(\theta)$,*

$$\mathrm{Var}(\zeta(k, f, h, g, \xi)) \leq c(\sigma(k,h)^2 + \tau(f,g)^2).$$

Proof: From (23) we infer

$$\mathrm{Var}(\zeta) \leq 2(\|w_1\|_{L_1(X)} + \|w_2\|_{L_1(X)})\|\varphi^2 p_0\|_{L_\infty(X)}.$$

By (25),

$$\|w_2\|_{L_1(X)} \leq \beta\tau(f,g)^2. \tag{33}$$

Now we estimate $\|w_1\|_{L_1(X)}$. According to (26),

$$\begin{aligned}
\|w_1\|_{L_1(X)} \leq\ & \|(I - T_{k^2p})^{-1} : L_1(X) \to L_1(X)\| \\
& \times \|T_{(k-h)^2p} + T_{(h^2-k^2)p}(I - T_{h^2p})^{-1}T_{(h-k)hp} : L_\infty(X) \to L_1(X)\| \\
& \times \|(I - T_{khp})^{-1} : L_\infty(X) \to L_\infty(X)\|\ \ \|f^2 q\|_{L_\infty(X)} \\
\leq\ & \beta\beta_0\theta(\|T_{(k-h)^2p} : L_\infty(X) \to L_1(X)\| \\
& + \|T_{(h^2-k^2)p}(I - T_{h^2p})^{-1}T_{(h-k)hp} : L_\infty(X) \to L_1(X)\|).
\end{aligned} \tag{34}$$

Obviously

$$\begin{aligned}
& \|T_{(k-h)^2p} : L_\infty(X) \to L_1(X)\| \\
\leq\ & \int_X \int_X (k(x,y) - h(x,y))^2 p(x,y)\, d\mu(x)\, d\mu(y) = \sigma(k,h)^2.
\end{aligned} \tag{35}$$

To estimate the second summand in (34), note first that given any $s \in L_\infty(X^2)$, we have, using Hölder's inequality,

$$\begin{aligned}
& \|T_{sp} : L_\infty(X) \to L_2(X)\|^2 \\
\leq\ & \int_X \left(\int_X |s(x,y)| p(x,y)\, d\mu(y) \right)^2 d\mu(x) \\
\leq\ & \int_X \left(\int_X s(x,y)^2 p(x,y)\, d\mu(y) \right) \left(\int_X p(x,y)\, d\mu(y) \right) d\mu(x) \\
\leq\ & \int_X \int_X s(x,y)^2 p(x,y)\, d\mu(x)\, d\mu(y)
\end{aligned} \tag{36}$$

and similarly, if s^* is defined by $s^*(x,y) = s(y,x)$, then

$$\begin{aligned}
& \|T_{sp} : L_2(X) \to L_1(X)\|^2 = \|T_{s^*p^*} : L_\infty(X) \to L_2(X)\|^2 \\
\leq\ & c_p \int_X \int_X s(x,y)^2 p(x,y)\, d\mu(x)\, d\mu(y).
\end{aligned}$$

It follows that

$$\begin{aligned}
& \|T_{(h^2-k^2)p}(I - T_{h^2p})^{-1}T_{(h-k)hp} : L_\infty(X) \to L_1(X)\|^2 \\
\leq\ & \|T_{(h^2-k^2)p} : L_2(X) \to L_1(X)\|^2 \|(I - T_{h^2p})^{-1} : L_2(X) \to L_2(X)\|^2 \\
& \times \|T_{(h-k)hp} : L_\infty(X) \to L_2(X)\|^2
\end{aligned}$$

$$\leq c_p\beta\beta_0 \int_X \int_X (h(x,y)-k(x,y))^2(h(x,y)+k(x,y))^2 p(x,y)\, d\mu(x)\, d\mu(y)$$
$$\times \int_X \int_X (h(x,y)-k(x,y))^2 h(x,y)^2 p(x,y)\, d\mu(x)\, d\mu(y)$$
$$\leq 4\alpha^4 c_p\beta\beta_0\sigma(k,h)^4. \tag{37}$$

4 Galerkin approximation

In this section we discuss one possible way of constructing h and g. We shall exploit the data supplied by a simple deterministic Galerkin method. Then we analyze this construction from the point of view of approximation quantities of section 3. So let us first consider the Galerkin scheme. We shall only deal with piecewise constant approximations. For this purpose fix $n \in \mathbb{N}$ and let $X = \cup_{i=1}^n X_i$ be a partition of X with $X_i \in \Sigma$, $\mu(X_i) \neq 0$ for all i and $\mu(X_i \cap X_j) = 0$ for all $i \neq j$. The Galerkin method for solving

$$u - T_K u = F$$

seeks to find an approximation to u of the form

$$\widetilde{u} = \sum_{i=1}^n u_i \chi_{X_i}$$

(χ denoting the characteristic function) satisfying

$$(\widetilde{u} - T_K\widetilde{u}\ ,\ \chi_{X_i}) = (F, \chi_{X_i}) \qquad (i = 1,\ldots,n).$$

This leads to the system

$$u_i - \sum_{j=1}^n K_{ij}u_j = F_i \qquad (i = 1,\ldots,n) \tag{38}$$

with

$$K_{ij} = \mu(X_i)^{-1} \int_{X_i}\int_{X_j} K(x,y)\, d\mu(y)\, d\mu(x) \tag{39}$$

$$F_i = \mu(X_i)^{-1} \int_{X_i} F(x)\, d\mu(x). \tag{40}$$

Often K_{ij} and F_i cannot be determined exactly, and one replaces them by some approximations, K'_{ij} and F'_i. Furthermore, in some situations the resulting system

$$u'_i - \sum_{j=1}^n K'_{ij}u'_j = F'_i \qquad (i = 1,\ldots,n) \tag{41}$$

itself is solved only approximately, by iterative methods. So let $(v_i)_{i=1}^n$ be an approximate solution of (41). We put

$$H_{ij} = \begin{cases} K'_{ij} & \text{if } (i,j) \in S \\ 0 & \text{otherwise,} \end{cases}$$

where $S \subset \{1,\ldots,n\}^2$ is supposed to be a large subset, whose choice allows us to pursue some strategy of excluding certain pairs of regions for which our construction might behave badly. We will comment on the choice of S later on. Finally, we compute $(G_i)_{i=1}^n$ by

$$v_i - \sum_{j=1}^{n} H_{ij} v_j = G_i \qquad (i = 1,\ldots,n). \tag{42}$$

This is our starting point - a system satisfied by $(v_i)_{i=1}^n$ exactly. Now we pose the following task. Find $h(x,y)$ on X^2 such that the following statement holds for all $(z_i)_{i=1}^n$, $(b_i)_{i=1}^n \in \mathbb{R}^n$: $(z_i)_{i=1}^n$ is a solution of the system

$$z_i - \sum_{j=1}^{n} H_{ij} z_j = b_i$$

iff $z = \sum_{i=1}^n z_i \chi_{X_i}$ is a solution of the integral equation in $L_\infty(X)$,

$$z - T_{hp} z = b,$$

with $b = \sum_{i=1}^n b_i \chi_{X_i}$. It is readily verified that this holds iff

$$(T_{hp} \chi_{X_j})(x) = H_{ij} \tag{43}$$

for almost all $x \in X_i$ and all $i,j = 1,\ldots,n$. (43) means that

$$\int_{X_j} h(x,y) p(x,y)\, d\mu(y) = H_{ij} \quad (x \in X_i, \quad i,j = 1,\ldots,n).$$

A possible choice results from assuming $h(x,y) = h_{ij}(x)$ for $x \in X_i$, $y \in X_j$, which gives

$$h_{ij}(x) = \begin{cases} K'_{ij}/p_j(x) & (i,j) \in S \\ 0 & (i,j) \notin S, \end{cases} \tag{44}$$

where

$$p_j(X) = \int_{X_j} p(x,y)\, d\mu(y). \tag{45}$$

We assume that S is chosen in such a way that for $(i,j) \in S$,

$$\operatorname*{ess\,inf}_{x \in X_i} p_j(x) > 0. \tag{46}$$

Then definition (44) is correct. We shall also assume

$$q(x) \geq q_0 > 0 \quad (x \in X). \tag{47}$$

Finally, we set

$$G = \sum_{i=1}^{n} G_i \chi_{X_i} \tag{48}$$

and

$$g(x) = G(x)/q(x). \tag{49}$$

Summarizing, we have found h and g in such a way that $v = \sum_{i=1}^{n} v_i \chi_{X_i}$ is the exact solution of

$$v - T_{hp} v = gq. \tag{50}$$

It is intuitively clear that the finer the Galerkin approximation and the larger S, the closer will T_{hp} be to T_{kp} and G be to F. The precise behaviour depends, of course, on the concrete situation considered. Nevertheless even in the general case it is possible to point out some quantities which on one hand determine the proximity of the neighboring equation to the original one, and on the other hand, are themselves determined by the precision of the Galerkin method. Thus our discussion aims at understanding how the variance reduction depends on the Galerkin quality. Let us introduce the following notation. Set

$$\bar{p}_{ij} = \mu(X_i)^{-1} \int_{X_i} \int_{X_j} p(x,y)\, d\mu(y)\, d\mu(x) = \mu(X_i)^{-1} \int_{X_i} p_j(x)\, d\mu(x) \tag{51}$$

and for $(i,j) \in S$,

$$\bar{k}_{ij} = K_{ij}/\bar{p}_{ij} \tag{52}$$

$$\bar{k}'_{ij} = K'_{ij}/\bar{p}_{ij}. \tag{53}$$

So $|K_{ij} - K'_{ij}|$ and $|\bar{k}_{ij} - \bar{k}'_{ij}|$ reflect the precision of approximating the true Galerkin coefficients (39).
Note that

$$\bar{k}_{ij} = \frac{\int_{X_i} \int_{X_j} k(x,y) p(x,y)\, d\mu(y)\, d\mu(x)}{\int_{X_i} \int_{X_j} p(x,y)\, d\mu(y)\, d\mu(x)},$$

so $\bar{k}_{ij}$ is an average of k over $X_i \times X_j$ with respect to a certain measure determined by p. Hence the quantity

$$\operatorname*{ess\,sup}_{x \in X_i, y \in X_j} |k(x,y) - \bar{k}_{ij}| = \kappa_{ij}$$

is related to continuity of k. Put furthermore

$$\operatorname*{ess\,inf}_{x\in X_i} p_j(x)/\bar{p}_{ij} = r_{ij}$$

for $(i,j) \in S$. According to (46), $r_{ij} > 0$ for $(i,j) \in S$. It is a goal of our strategy S to keep r_{ij} reasonably away from zero. This quantity together with the following,

$$\operatorname*{ess\,sup}_{x\in X_i} \frac{|p_j(x) - \bar{p}_{ij}|}{\bar{p}_{ij}} = \lambda_{ij},$$

also defined for $(i,j) \in S$, describes the oscillation of $p_j(x)$ on elements X_i. The finer the partition, the closer r_{ij} to 1 and the smaller λ_{ij} - as a rule. Put finally for all $i,j = 1,\ldots,n$,

$$\pi_{ij} = \mu(X_i)\bar{p}_{ij} = \int_{X_i}\int_{X_j} p(x,y)\; d\mu(y)\; d\mu(x).$$

It follows that

$$\sum_{i,j=1}^{n} \pi_{ij} = \int_X (1 - q(x))\; d\mu(x) \le \mu(X).$$

The following two propositions show how the quantities which control the variance reduction are related to those of the Galerkin approximation discussed above.

Proposition 6 *If $k \in L_\infty(X^2)$ with $\|k\|_{L_\infty(X^2)} \le \alpha$, then*

$$\sigma(k,h)^2 \le \; 3\textstyle\sum_{(i,j)\in S}(\kappa_{ij}^2 + \alpha^2 r_{ij}^{-1}\lambda_{ij}^2 + r_{ij}^{-1}(\bar{k}_{ij} - \bar{k}'_{ij})^2)\pi_{ij} + \alpha^2 \sum_{(i,j)\notin S} \pi_{ij}.$$

Proof:

$$\begin{aligned}
\sigma(k,h)^2 &= \int_X\int_X (k(x,y) - h(x,y))^2 p(x,y)\; d\mu(y)\; d\mu(x) \\
&= \sum_{(i,j)\in S} \int_{X_i}\int_{X_j} (k(x,y) - K'_{ij}/p_j(x))^2 p(x,y)\; d\mu(y)\; d\mu(x) \\
&\quad + \sum_{(i,j)\notin S} \int_{X_i}\int_{X_j} k(x,y)^2 p(x,y)\; d\mu(y)\; d\mu(x). \qquad (54)
\end{aligned}$$

For $(i,j) \in S$, we have, taking into account (53),

$$\begin{aligned}
&\int_{X_i}\int_{X_j} (k(x,y) - K'_{ij}/p_j(x))^2 p(x,y)\; d\mu(y)\; d\mu(x) \\
\le\; &3\int_{X_i}\int_{X_j} (k(x,y) - \bar{k}_{ij})^2 p(x,y)\; d\mu(y)\; d\mu(x)
\end{aligned}$$

$$
\begin{aligned}
&+3\int_{X_i}\int_{X_j}(\bar{k}_{ij}-\bar{k}_{ij}\bar{p}_{ij}/p_j(x))^2 p(x,y)\ d\mu(y)\ d\mu(x)\\
&+3\int_{X_i}\int_{X_j}((\bar{k}_{ij}-\bar{k}'_{ij})\bar{p}_{ij}/p_j(x))^2 p(x,y)\ d\mu(y)\ d\mu(x)\\
\leq\ & 3\kappa_{ij}^2\pi_{ij}+3\bar{k}_{ij}^2\int_{X_i}(p_j(x)-\bar{p}_{ij})^2 p_j(x)^{-1}\ d\mu(x)+3(\bar{k}_{ij}-\bar{k}'_{ij})^2\int_{X_i}\bar{p}_{ij}^2 p_{ij}(x)^{-1}\ d\mu(x)\\
\leq\ & 3\kappa_{ij}^2\pi_{ij}+3\bar{k}_{ij}^2 r_{ij}^{-1}\lambda_{ij}^2\bar{p}_{ij}\mu(X_i)+3(\bar{k}_{ij}-\bar{k}'_{ij})^2 r_{ij}^{-1}\bar{p}_{ij}\mu(X_i)\\
\leq\ & 3(\kappa_{ij}^2+\alpha^2 r_{ij}^{-1}\lambda_{ij}^2+r_{ij}^{-1}(\bar{k}_{ij}-\bar{k}'_{ij})^2)\pi_{ij},
\end{aligned}
$$

which proves the proposition.

Proposition 7 *If $k\in L_\infty(X^2)$ with $\|k\|_{L_\infty(X^2)}\leq\alpha$, then*

$$\tau(f,g)^2\leq 2q_0^{-1}((1+\alpha)^2\mu(X)\|u-v\|_{L_\infty(X)}^2+\sigma(k,h)^2\|v\|_{L_\infty(X)}^2).$$

Proof:

$$
\begin{aligned}
\tau(f,g)^2 &= \int_X (F(x)-G(x))^2 q(x)^{-1}\ d\mu(x)\\
&\leq q_0^{-1}\|F-G\|_{L_2(X)}^2\\
&= q_0^{-1}\|(I-T_{kp})u-(I-T_{hp})v\|_{L_2(X)}^2\\
&\leq 2q_0^{-1}(\|(I-T_{kp})(u-v)\|_{L_2(X)}^2+\|(T_{hp}-T_{kp})v\|_{L_2(X)}^2)\\
&\leq 2q_0^{-1}((\|I-T_{kp}:L_\infty(X)\to L_2(X)\|)^2\|u-v\|_{L_\infty(X)}^2\\
&\quad +\|T_{(h-k)p}:L_\infty(X)\to L_2(X)\|^2\|v\|_{L_\infty(X)}^2)\\
&\leq 2q_0^{-1}((\mu(X)^{1/2}+\alpha\mu(X)^{1/2})^2\|u-v\|_{L_\infty(X)}^2+\sigma(k,h)^2\|v\|_{L_\infty(X)}^2),
\end{aligned}
$$

where we used relation (36) twice. This proves the desired inequality.

The last proposition shows the direct dependence of $\tau(f,g)$ on $\|u-v\|_{L_\infty(X)}$, which is the error of the Galerkin solution, and on $\sigma(k,h)$, which was already estimated above.

Now suppose that $k\in\mathcal{K}(\alpha,\gamma,n_0)$ and h is constructed as above. Can we assert that h also belongs to such class, at least asymptotically? In fact, under certain assumptions, we can. Let $\alpha'>\alpha$ and $\gamma<\gamma'<1$. In a way similar to the proof of Proposition 6 we can derive that for $(i,j)\in\mathcal{S}$,

$$\operatorname*{ess\,sup}_{x\in X_i,y\in X_j}|k(x,y)-h(x,y)|\leq\kappa_{ij}+\bar{k}_{ij}r_{ij}^{-1}\lambda_{ij}+|\bar{k}_{ij}-\bar{k}'_{ij}|r_{ij}^{-1}.$$

Let us assume that the quantities on the right hand side are small, giving that

$$\max_{(i,j)\in S} \operatorname*{ess\,sup}_{x\in X_i, y\in X_j} |k(x,y) - h(x,y)| = \delta.$$

Since $h(x,y) = 0$ for $x \in X_i$, $y \in X_j$ and $(i,j) \notin S$, we get

$$|h(x,y)| \leq |k(x,y)| + \delta$$

for almost all $x, y \in X$. By the definition of $\mathcal{K}(\alpha, \gamma, n_o)$,

$$\|T^{n_0}_{k^2p} : L_\infty(X) \to L_\infty(X)\| \leq \gamma.$$

A perturbation argument gives now

$$\|T^{n_0}_{h^2p} : L_\infty(X) \to L_\infty(X)\| \leq \gamma'$$

provided δ is small enough. This argument carries over to the case of $\mathcal{K}'$, provided k satisfies, in addition,

$$\|T^{n_1}_{k^2p} : L_1(X) \to L_1(X)\| \leq \gamma_1 < 1.$$

for some n_1. This is the case e. g. for the radiosity equation described in the next section.

5 Application to the radiosity equation

In this section we shall illustrate the method developed above by applying it to the radiosity equation of computer graphics [CW93]. This equation describes the illumination of closed scenes (interior of buildings, rooms with furniture etc.). The same type of equation also occurs in radiative heat transfer [SH92]. The radiosity equation has the form

$$u(x) = \int_X K(x,y)u(y)\, d\mu + F(x) \tag{55}$$

where X is the surface of the scene, $F(x)$ is the radiometric intensity of light sources at x, and $u(x)$ is the total intensity of light due to the contribution of sources and (multiple) reflection. Only monochromatic light is considered, the full colour image has to be obtained by superposition [CW93]. It is assumed that only diffuse reflections take place (i. e. no mirrors, gloss, etc.). Then the kernel is given by

$$K(x,y) = \rho(x)w(x,y)\frac{\cos(n_x, y-x)\cos(n_y, x-y)}{\pi|x-y|^2}.$$

Here n_x is the surface normal in point x (directed into the scene), $\cos(n_x, y-x)$ denotes the cosine of the angle between the vectors n_x and $y-x$, and $w(x,y)$ is the visibility factor - it is equal to 1 if the line segment from x to y does not intersect the scene in

any other point (x and y "see each other "), and equal to 0 otherwise. The factor $\rho(x)$ describes the portion of the incoming energy which is reflected (the rest is absorbed). In diffuse environments one assumes $\rho(x) < 1$ for all x. For details on the radiosity equation we refer to [CW93]. Often scence are composed of triangles (or other elementary geometric objects). We shall assume that $X = \cup_{i=1}^{n} X_i$ is some decomposition into triangles (possibly not the original, but already refined ones), satisfying $\mu(X_i \cap X_j) = 0$, where μ is the Lebesgue surface measure.

A standard deterministic approach of solving (55) is the radiosity method. This is just the Galerkin scheme with piecewise constant approximation on X_i as described in section 4. It is assumed that $\rho(x) \equiv \rho_i$ is constant on X_i, and $F(x)$ is either also assumed to be constant or replaced by the average emittance

$$F_i = \mu(X_i)^{-1} \int_{X_i} F(x)\, d\mu.$$

Then we get, following the notation of section 4,

$$K_{ij} = \rho_i \Psi_{ij}$$

with

$$\Psi_{ij} = \mu(X_i)^{-1} \int_{X_i} \int_{X_j} Q(x,y)\, dy\, dx$$

and

$$Q(x,y) = w(x,y) \frac{\cos(n_x, y-x)\cos(n_y, x-y)}{\pi |x-y|^2}.$$

The quantity Ψ_{ij} is the so-called (patch-to-patch) form factor ([CW93], ch. 4). The resulting system (38), (or (41), when only approximations to the form factors are available) is called the radiosity system. Usually it is solved iteratively (radiosity methods).

Let us now turn to Monte Carlo methods of solving (55). They are applied when, due to the complexity of the scene, the Galerkin method is too time-consuming or not precise enough. One simulates either directly the light transfer ("from the source to the eye") or the "importance" transfer based on the dual equation ("from the eye to the source"). The latter corresponds to the scheme described in section 2, and we will restrict our considerations to this case. Now suppose we seek the value of u in a point x_0. Then the initial distribution of the Markov process is the delta function in x_0 (compare remark 2 after Lemma 1). We fix the transition density as

$$p(x,y) = (1 - q_0)\, Q(x,y).$$

$Q(x,y)$ is the cosine distribution over the hemisphere in x of directions into the scene, and $0 < q_0 < 1$ a termination parameter at our disposal. (There are many modifications

to produce random walks through the scene, but for our purposes of illustrating the developed method we consider only this one.)

In the notation of section 2 we have $\varphi \equiv 1$, $k(x,y) = \rho_i(1-q_0)^{-1}$ for $x \in X_i$, $y \in X$, and $f(x) = F_i/q_0$ for $x \in X_i$. So the walk (6) and its contribution (7) are defined. (Note that if the light sources occupy only a small portion of the scene, the last step of the walk should be modified in such a way that x_m always hits a light source. This also requires a corresponding modification of the last step of our method.)

Now let us discuss the scheme of section 4 for the present situation. Having fixed S, we define $h_{ij}(x)$ by (44), with

$$p_j(x) = (1-q_0)\int_{X_j} Q(x,y)\,dy.$$

The integral is called the differential (point-to-patch), form factor ([CW93], ch. 4). If X_j is fully visible from x, i. e. $w(x,y) = 1$ for all $y \in X_j$, then this form factor can be computed explicitly ([CW93], ch. 4.6). Hence our strategy should ensure $w(x,y) \equiv 1$ for $x \in X_i$, $y \in X_j$, $(i,j) \in S$. As a rule, the finer the partition, the more pairs of elements X_i, X_j will satisfy $w(x,y) \equiv 1$ or $w(x,y) \equiv 0$ on $X_i \times X_j$. Next observe that

$$\begin{aligned} \bar{p}_{ij} &= (1-q_0)\ \mu(X_i)^{-1}\int_{X_i}\int_{X_j} Q(x,y)\,dy\,dx \\ &= (1-q_0)\Psi_{ij}, \end{aligned}$$

with Ψ_{ij} the form factor, as above. The discussion in section 4 made it clear that the strategy should furthermore exclude those (i,j) for which the differential form factor deviates to much from its mean - the form factor. This is essentially the case when the view of X_j from x "collapses" as x moves through X_i (i. e. $p_j(x) = 0$ for some $x \in X_i$). The way of setting up S should be investigated in more detail.

In the rest of this section we want to treat a numerical example. We move from three-dimensional graphics to dimension two ("flatland", see e. g. [Hec92]). Our "scene" is now the unit circle, $X = \{e^{is} : 0 \le s \le 2\pi\}$, and the radiosity equation takes the form

$$u(x) = F(x) + \rho(x)\int_X \frac{\cos(n_x, y-x)\cos(n_y, x-y)}{2|x-y|}u(y)\,dy.$$

Using the parametrization above, and taking into account that for the circle, $|x-y| = 2\cos(n_y, x-y)$, we get

$$u(s) = F(s) + \frac{\rho(s)}{4}\int_{s-\pi}^{s+\pi} \cos(\pi/2 - |t-s|/2)u(t)\,dt,$$

thus

$$u(s) = F(s) + \frac{\rho(s)}{4}\int_0^{2\pi} |\sin((t-s)/2)|u(t)\,dt. \tag{56}$$

This is the radiosity equation for the circle. Elementary calculations verify that for $C, D \in \mathbb{R}$, $\ell \in \mathbb{N}$ and

$$F(s) = C(1-\rho(s)) + D(1+(2\ell-1)^{-1}(2\ell+1)^{-1}\rho(s))\sin \ell s, \tag{57}$$

the exact solution of (56) is

$$u(s) = C + D\sin \ell s. \tag{58}$$

In our experiments we compare three algorithms - the radiosity method (i. e. the Galerkin scheme), the standard Monte Carlo method, and the scheme of section 4, which we shall call preconditioned Monte Carlo. We shall solve equation (56) with piecewise constant $\rho(s)$ over an initial equidistant partition of $[0, 2\pi]$ into n_0 intervals. We fix ρ^* with $0 < \rho^* < 1$ and set

$$\rho(s) = \rho_i \qquad \left(s \in \left[\tfrac{2\pi(i-1)}{n_0}, \tfrac{2\pi i}{n_0}\right), \quad i = 1, \ldots, n_0\right),$$

where the ρ_i are chosen randomly on $[0, \rho^*]$. Furthermore, we choose at random m_0 points on $[0, 2\pi]$, in which the values of the solution are supposed to be computed. The ρ_i's and the points remain fixed throughout all experiments.

The Galerkin scheme is built on an equidistant partition of $[0, 2\pi]$ into n intervals, with $n = n_0 n_1$, $n_1 \in \mathbb{N}$. The coefficients and the right-hand sides of the resulting system can be computed explicitly. It is solved by Gauss-Seidel iteration, with an error, negligible over the discretization error.

The standard Monte Carlo scheme starts in the point in which the value has to be computed. Then it moves according to the transition density

$$p(x, y) = (1-q_0)\frac{\cos(n_x, y-x)\cos(n_y, x-y)}{2|x-y|}$$

or, in the parametrization,

$$p(s,t) = \frac{1-q_0}{4}|\sin\frac{t-s}{2}|,$$

which can easily be generated by inversion. This leads to

$$k(s,t) = (1-q_0)^{-1}\rho_i \qquad \left(s \in \left[\tfrac{2\pi(i-1)}{n_0}, \tfrac{2\pi i}{n_0}\right), \quad t \in [0, 2\pi]\right).$$

The ingredients $p_j(x)$ of the improved Monte Carlo method become

$$p_j(s) = (1-q_0)\sin\left(\frac{(2j-1)\pi}{2n} - \frac{s}{2}\right)\sin\frac{\pi}{2n},$$

for $s \notin [\frac{2\pi(j-1)}{n}, \frac{2\pi j}{n})$, and we set

$$S = \{(i,j): \quad i, j = 1, \ldots, n, \quad i \neq j\}.$$

For our experiments we put $n_0 = m_0 = 10$, $\rho^* = 0.75$, $q_0 = 0.2$, $C = 5$, $D = 1$, $\ell = 5$ (this guarantees $F(x) \geq 0$ and $u(x) \geq 0$ for all x). For each of the three methods (radiosity, standard Monte Carlo, preconditioned Monte Carlo) we compute the absolute error, for the two stochastic methods also the square root of the empirical variance. To give a more conclusive impression of the three processes, we average each of the quantities over the ten points in which the solution is computed. (We also computed the worst case and observed the same tendencies as in the average case.) This is carried out for various n (size of the Galerkin system) and N (number of realizations of the Markov chain).

The Gauss-Seidel method was performed with 10 iteration (we observed that already at five iterations the precision matching that of the table below was reached). Note that the cost of the Galerkin method is $O(n^2)$, while that of standard Monte Carlo is $O(N)$, so a balanced choice would be $N = O(n^2)$. Nevertheless we included also other pairs to show the relative behaviour. The computations were done on an HP 9000/735/99 workstation, in double precision. The results are given in the following table

n	N	e^{rad}	e^{mc}	e^{pmc}	$\sqrt{\text{Var}^{\text{mc}}}$	$\sqrt{\text{Var}^{\text{pmc}}}$
10	10	.286E+0	.126E+1	.139E+0	.776E+1	.621E+0
	100	.286E+0	.644E+0	.398E−1	.714E+1	.659E+0
	1000	.286E+0	.251E+0	.298E−1	.777E+1	.717E+0
	10000	.286E+0	.655E−1	.208E−1	.770E+1	.693E+0
	100000	.286E+0	.233E−1	.183E−1	.767E+1	.693E+0
100	10	.364E−1	.126E+1	.150E−1	.776E+1	.764E−1
	100	.364E−1	.644E+0	.442E−2	.714E+1	.891E−1
	1000	.364E−1	.251E+0	.284E−2	.777E+1	.955E−1
	10000	.364E−1	.655E−1	.502E−3	.770E+1	.983E−1
	100000	.364E−1	.233E−1	.371E−3	.767E+1	.980E−1
1000	10	.344E−2	.126E+1	.973E−3	.776E+1	.704E−2
	100	.344E−2	.644E+0	.349E−3	.714E+1	.803E−2
	1000	.344E−2	.251E+0	.286E−3	.777E+1	.844E−2
	10000	.344E−2	.655E−1	.781E−4	.770E+1	.906E−2
	100000	.344E−2	.233E−1	.190E−4	.767E+1	.874E−2
	1000000	.344E−2	.720E−2	.660E−5	.768E+1	.880E−2

Here $e^{\text{rad}}, e^{\text{mc}}$ and e^{pmc} are respectively the averaged absolute errors of the radiosity (i. e. Galerkin), standard Monte Carlo and preconditioned Monte Carlo method, while $\sqrt{\text{Var}^{\text{mc}}}$ and $\sqrt{\text{Var}^{\text{pmc}}}$ denote the square roots of the the empirical variances of the respective processes.

References

[CW93] M. F. Cohen and J. R. Wallace. *Radiosity and Realistic Image Synthesis.* Academic Press, 1993.

[DS58] N. Dunford and J. T. Schwartz. *Linear Operators. Part I: General Theory.* Interscience Publishers, New York, London, 1958.

[EM82] S. M. Ermakov and G. A. Mikhailov. *The Theory of Statistical Trials* (in Russian). Nauka, Moscow, 1982.

[ENS89] S. M. Ermakov, W. W. Nekrutkin, and A. S. Sipin. *Random Processes for Classical Equations of Mathematical Physics*, volume 34 of Math. Appl. (Soviet Ser.). Kluwer, Dodrecht, 1989.

[Erm71] S. M. Ermakov. *The Monte Carlo Method and Related Problems* (in Russian). Nauka, Moscow, 1971.

[ES85] S. M. Ermakov and A. S. Sipin. A new scheme of the Monte Carlo method for solving problems of mathematical physics. *Soviet Phys. Doklady*, **30**(11): 935 – 936. 1985.

[Hec92] P. Heckbert. Radiosity in flatland. *Computer Graphics Forum (Eurographics '92)*, **11**: 181 – 192, 1992.

[Hei94] S. Heinrich. Random approximation in numerical analysis. In K. D. Bierstedt, A. Pietsch, W. M. Ruess, and D. Vogt, editors, *Functional Analysis*, pages 123 – 171. Marcel Dekker, New York, 1994.

[HM93] S. Heinrich and P. Mathé. The Monte Carlo complexity of Fredholm integral equations. *Math. of Computation*, **60**: 257 – 278, 1993.

[KW86] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods, Vol. I: Basics.* Wiley, New York, 1986.

[LW94] E. P. Lafortune and Y. D. Willems. The ambient term as a variance reducing technique for Monte Carlo ray tracing. In *Proc. 5th Eurographics Workshop.* Darmstadt, 1994.

[Mat] P. Mathé. Approximation Theory of Stochastic Numerical Methods. to appear.

[Mik91a] G. A. Mikhailov. *Minimization of Computational Costs of Non-Analogue Monte Carlo Methods.* World Scientific, Singapore, 1991.

[Mik91b] G. A. Mikhailov. *Optimization of Weighted Monte Carlo Methods.* Springer, 1991.

[Nov88] E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis.* Lecture Notes in Mathematics 1349. Springer: Berlin, Heidelberg, New York, London, Paris and Tokio, 1988.

[Sab91] K. K. Sabelfeld. *Monte Carlo Methods in Boundary Value Problems.* Springer, Berlin-Heidelberg-New York, 1991.

[SG69] J. Spanier and E. M. Gelbard. *Monte Carlo Principles and Neutron Transport Problems.* Addison-Wesley, Reading, Massachusetts, 1969.

[SH92] R. Siegel and J. R. Howell. *Thermal Radiation Heat Transfer.* Hemisphere Publishing Corporation, New York, 3rd edition, 1992.

[Sob73] I. M. Sobol. *Computational Monte Carlo Methods* (in Russian). Nauka, Moscow, 1973.

[Spa79] J. Spanier. A new family of estimators for random walk problems. *J. Inst. Math. Appl.*, **23**: 1 – 31, 1979.

[Tri78] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators.* Deutscher Verlag der Wissenschaften, Berlin, 1978.

[TWW88] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity.* Academic Press, New York, 1988.