Le contrôle de l'erreur dans la méthode de radiosité hiérarchique

Nicolas Holzschuch

Le contrôle de l'erreur dans la méthode de radiosité hiérarchique

Nicolas HOLZSCHUCH

Thèse présentée pour l'obtention du titre de : Docteur de l'Université Joseph Fourier – Grenoble I, spécialité Informatique (arrêtés ministériels du 5 juillet 1984 et du 30 mars 1992), soutenue le 5 mars 1996.

Composition du jury :

Jacques VOIRON, professeur à l'Université Joseph Fourier, président ;
Didier ARQUÈS, professeur à l'Université Marne-la-Vallée, rapporteur ;
Frederik W. JANSEN, professeur à la Technische Universiteit Delft, rapporteur ;
Bernard PÉROCHE, professeur à l'École des Mines de Saint-Étienne ;
Claude PUECH, professeur à l'Université Joseph Fourier, directeur ;
François SILLION, chargé de recherche au CNRS, HDR.

Thèse préparée au sein du laboratoire *i*MAGIS/IMAG. (*i*MAGIS est un projet commun entre le CNRS, l'INRIA, l'INPG et l'UJF).

À Myriam

Résumé

Nous présentons ici plusieurs améliorations d'un algorithme de modélisation de l'éclairage, la méthode de radiosité. Pour commencer, une analyse détaillée de la méthode de radiosité hiérarchique permet de souligner ses points faibles et de mettre en évidence deux améliorations simples : une évaluation paresseuse des interactions entre les objets, et un nouveau critère de raffinement qui élimine en grande partie les raffinements inutiles. Un bref rappel des propriétés des fonctions de plusieurs variables et de leurs dérivées suit, qui permet d'abord de déduire une réécriture de l'expression de la radiosité, d'où un calcul numérique plus précis. Les méthodes d'estimation de l'erreur produite au cours du processus de modélisation de la lumière sont introduites. Nous voyons alors comment les propriétés de concavité de la fonction de radiosité permettent – grâce au calcul des dérivées successives de la radiosité – un contrôle complet de l'erreur commise dans la modélisation des interactions entre les objets, et donc un encadrement précis de la radiosité. Nous présentons un critère de raffinement basé sur cette modélisation des interactions, et un algorithme complet de radiosité hiérarchique intégrant ce critère de raffinement, et donc permettant un contrôle de l'erreur commise sur la radiosité au cours de la résolution. Finalement, nous présentons les méthodes de calcul pratique des dérivées successives de la radiosité (gradient et Hessien) dans le cas d'un émetteur constant sans obstacles tout d'abord, puis dans le cas d'un émetteur constant en présence d'obstacles et dans le cas d'un émetteur sur lequel la radiosité varie de façon linéaire.

Mots clef : Images de synthèse, simulation physique, radiosité, radiosité hiérarchique, analyse d'algorithmes, facteurs de forme, contrôle de l'erreur, dérivées de la radiosité, visibilité, évaluation paresseuse, fonctions de base d'ordre supérieur.

Abstract

We introduce several improvements to the hierarchical radiosity method. First, a complete analysis of a specific implementation of the hierarchical radiosity method allows to point out its bottlenecks. Based on this analysis, we suggest two simple improvements: a lazy evaluation of top-level interactions, and a new refinement criterion, that greatly reduces the number of interactions, without loss of precision. A brief introduction to the properties of functions of several variables and their derivatives follows, which allows a rewriting of the expression of radiosity, and hence a better numerical approximation. Methods for the estimation of the error produced during the radiosity computations are analysed. We then introduce the concavity properties of the radiosity function that, combined with an exact computation of the radiosity derivatives, allow a complete control of the error on the interactions between patches, and hence a precise minoration and majoration of the radiosity on all the patches. We introduce a new refinement criterion based on this modelling of interactions, and a complete hierarchical radiosity algorithm using this refinement criterion. The last part of the thesis is devoted to practical computations of the radiosity derivatives (gradient and Hessian), first for a constant emitter with total visibility, then for a constant emitter with partial visibility and for an emitter with linear radiosity.

Keywords: Computer Graphics, Image Synthesis, Physical Simulation, Radiosity, Hierarchical Radiosity, Form-Factor Computation, Error Control, Radiosity Derivatives, Profiling, Visibility, Lazy Linking, Higher Order Basis Functions.

Table des matières

Table des figures

Liste des tableaux

Remerciements

Je tiens tout d'abord à remercier M. Jacques Voiron, qui me fait l'honneur de présider ce jury. Ma gratitude va également à M. Didier Arquès et M. Frederik Jansen, qui ont accepté d'être les rapporteurs de cette thèse, et qui ont accompli ce travail dans des délais très courts, malgré leur emploi du temps chargé. Je remercie également M. Bernard Péroche qui a bien voulu faire partie de ce jury de thèse et dont j'ai apprécié à diverses reprises le contact stimulant.

Tout au long de mes recherches, Claude Puech a montré une étonnante capacité d'analyse rapide et de critique constructive de mes orientations de travail : ses conseils ont toujours été d'une pertinence et d'une acuité extrêmes, montrant sa complète compréhension du sujet, de ses intérêts et de ses points faibles. Sa connaissance des arcanes de l'enseignement supérieur a été précieuse pour le financement de cette thèse. Pour tout cela, je le remercie.

François Sillion a été pour moi à la fois un directeur de recherches et un exemple de rigueur intellectuelle. Je tiens à le remercier pour ses conseils, qui m'ont évité de me disperser au cours de la thèse, et pour sa disponibilité constante, malgré un emploi du temps familial et professionnel chargé. Sa grande expérience, dans le domaine de l'informatique comme dans celui des sciences physiques, lui a permis de nombreuses fois de résoudre d'un mot un problème qui m'avait occupé pendant des heures.

Depuis son arrivée dans l'équipe *i*MAGIS, George Drettakis a assuré, lors de nombreuses discussions constructives, un rôle de conseil à la fois efficace et stimulant. J'ai notamment apprécié son étonnante capacité à repérer une piste de recherches prometteuse au sein d'une explication éventuellement embrouillée, et sa parfaite connaissance de tout ce qui concerne la modélisation de l'éclairage.

Une partie de ce travail a été déclenchée par les conversations que j'ai pu avoir lors de mon séjour au Fraunhofer Institut de Darmstadt avec Stefan Müller et les membres de son équipe : Wolfram Kresse, Dirk Reiners, Mathias Unbescheiden et Colette El Cacho. Je les remercie ici pour m'avoir donné une vision plus claire des besoins en matière d'utilisation industrielle des méthodes de modélisation de l'éclairage.

La rare qualité de l'ambiance qui règne au sein de l'équipe *i*MAGIS, les échanges permanents entre les membres de l'équipe, et la solidarité manifeste ont été précieuses tout au long des années. Outre un soutien moral constant et apprécié, ils m'ont permis de connaître les algorithmes utilisés dans le graphisme et l'image de synthèse. Si je n'ai pas utilisé de surfaces implicites, ce n'est pas faute d'avoir essayé !

Plus particulièrement, je remercie Dominique Gascuel pour son encadrement et ses conseils en matière de gestion du système ; Alexis Lamouret pour son amical voisinage

depuis le DEA ; Frédéric Durand pour sa constante disponibilité, sa gentillesse, son humour et pour les nâns au fromage ; Mathieu Desbrun sans la générosité et la disponibilité de qui l'ambiance au sein de l'équipe ne serait pas ce qu'elle est ; et François Faure pour son amitié.

Je tiens à rendre hommage à mes anciens maîtres, André Warusfel, Jean-Pierre Sarmant et André Jacquelin, pour m'avoir communiqué, outre la connaissance de leur matière, des exemples à la fois de rigueur et d'enthousiasme. Cet enthousiasme pour l'enseignement et la recherche m'a influencé durablement, notamment dans mes choix au cours de mes études.

Ma gratitude va aussi à tous ceux qui m'ont soutenu de leur amitié et aidé depuis de nombreuses années : Jacques Garrigue (ETA), Erwan David (Carlos), Loïc Grenié (Surjector), Marc Herzlich, Alain Gounon et Karim Belabas (KB), Étienne de la Vaissière, Olivia Mailly, Pascale Brillet et Marie-Catherine Olivi. Depuis notre installation à Grenoble, j'ai beaucoup apprécié le soutien, l'amicale présence et la solidarité de Mathias Houssay, de Mathieu Savin et Marie-Laure Leroy, de Sébastien Bornot et des moniteurs du CIES.

Je veux remercier tous ceux qui ont été mes étudiants au cours des quatre dernières années, et qui ont supporté avec patience et humour les maladresses et les distractions d'un débutant. Tous, à Évry, à Marne-la-Vallée et à Grenoble, ils m'ont énormément apporté, et mon seul regret est de ne pouvoir les remercier tous ici nommément.

Ce que j'ai appris et reçu d'Antoine Chambert-Loir est immense. Outre son enthousiasme pour toutes choses, ses qualités humaines et son immense culture pluridisciplinaire, je n'oublie pas sa disponibilité et ses conseils, directs et pertinents. Surtout, c'est lui qui m'a enseigné la grande valeur de l'amitié.

Christophe «Plume» Bonnet est un des dernier spécimens d'humaniste qu'il m'ait été donné de rencontrer. Sa maîtrise parfaite des subtilités de la langue française, alliée à une connaissance de la science informatique et de tous ses raffinements m'ont toujours impressionné. Mais ce que je retiens surtout de lui, c'est sa capacité à deviner à l'avance les endroits où l'on a besoin d'aide, et à apporter un soutien discret autant qu'efficace, et la force de son amitié.

Ce travail doit beaucoup à mes parents, que je veux remercier ici pour tout leur soutien affectif, pour la liberté de choix qu'ils m'ont laissé au cours de mes études et pour l'exemple qu'ils représentent pour moi. Je n'oublie pas non plus tout ce que je dois au soutien fraternel et à la bonne humeur constante d'Elisabeth, de Claire et de Marie-Laure.

Et puis... et puis il y a celle dont je ne parlerai pas parce que c'est personnel, celle qui a pourtant tout fait dans ce travail de recherche, celle qui a partagé tous les instants de ma thèse, et qui a fait bien plus encore : Myriam.

1.
Introduction

> Toute la conduite de notre vie dépend de nos sens, entre lesquels celui de la vue étant le plus universel et le plus noble, il n'y a point de doute que les inventions qui servent à augmenter sa puissance ne soient des plus utiles qui puissent être.
>
> René DESCARTES, *La Dioptrique*, 1637.

DANS son introduction au discours premier de *La dioptrique*, *De la lumière* [8], Descartes avait en vue les inventions récentes de son temps, notamment les lunettes d'approche.

L'idée même de pouvoir calculer artificiellement l'image d'une scène donnée à partir d'une modélisation des propriétés géométriques lui était totalement étrangère. Cependant, ce même discours premier contient déjà une modélisation de propriétés de réflectivité des corps, à la fois pour les réflecteurs «polis» – nous disons aujourd'hui spéculaires – que pour «les corps qui sont colorés et non polis », ceux que nous appelons aujourd'hui lambertiens.

Mais si la modélisation des propriétés de la lumière et des corps réfléchissants peut être remontée aussi loin que le XVII^{e} siècle, l'utilisation effective de ces même propriétés est plus récente. En 1936, Moon et Parry [30] calculent l'intensité de l'éclairage en différents points d'une pièce, pour vérifier s'il est possible d'y travailler dans de bonnes conditions.

Plus proche de nous, l'utilisation de logiciels de tracés d'ombres, par des urbanistes dans un but de planification des constructions en milieu urbain commence au début des années 70, avec l'algorithme scan-line (voir Appel, 1968 [1]), l'ancêtre du lancer de rayons. La recherche suivie dans le domaine de la synthèse d'image depuis cette date a permis dans un premier temps d'augmenter la qualité visuelle des images produites. Ce n'est que plus récemment, avec les travaux de Goral, Torrance et Greenberg, en 1984 [15], qu'apparaît la recherche du réalisme *physique*, la précision dans la modélisation des propriétés physiques de la lumière.

Les applications industrielles de la synthèse d'images se développent à mesure que les développements de la recherche rendent la technique accessible. Par exemple, le Fraunhofer Institut de Darmstadt est chargé de modéliser le futur aéroport de Francfort avant sa construction, pour permettre au maître d'œuvre de vérifier visuellement les travaux des architectes, et notamment le confort – en termes d'éclairage – des postes de travail. De la même manière, il est courant pour les compagnies de radiotéléphone de

modéliser la diffusion des ondes du radiotéléphone en milieu urbain, afin de pouvoir minimiser le nombre de points de réémission. Dans de telles applications industrielles, il est important de pouvoir contrôler l'erreur commise au cours du processus de simulation et de modélisation de la lumière.

La méthode de radiosité, parce qu'elle modélise des interactions entre éléments de surfaces au lieu d'effectuer des calculs ponctuellement rend possible une modélisation globale de l'ensemble des interactions, et donc un contrôle de l'erreur sur ces interactions. La complexité quadratique de la méthode de radiosité originale rend un tel contrôle très coûteux, et donc impossible en pratique. En revanche, la méthode de radiosité hiérarchique, due à Hanrahan ([16, 17]), et dont la complexité est linéaire par rapport au nombre d'interactions, rend possible un contrôle de l'erreur commise au cours du processus de simulation de l'éclairage.

Ce contrôle de l'erreur passe par une modélisation précise des interactions entre les objets composant la scène. Cette modélisation des interactions avec calcul de l'erreur commise est l'un des défis les plus stimulants pour la recherche dans les années à venir. Des travaux antérieurs ont montré qu'un contrôle précis de l'erreur commise, parce qu'il permet d'économiser les ressources système, en concentrant les calculs sur les points vraiment délicats, aboutit sur des scènes complexes à un gain de temps notable (voir Lischinski et al., [26]).

Le contrôle de l'erreur au cours de la simulation de l'éclairage passe par un contrôle de l'erreur commise sur chaque interaction. Ce contrôle ne peut se faire uniquement à partir des valeurs de la radiosité que l'on sait, par ailleurs, calculer avec précision en chaque point. La connaissance des dérivées successives de la radiosité est un outil fondamental pour le contrôle de l'erreur.

Nous verrons d'abord en quoi la méthode de radiosité, et en particulier la méthode de radiosité hiérarchique se distingue des autres algorithmes de synthèse d'images. Les différentes évolutions de l'algorithme de radiosité hiérarchique depuis son introduction en 1990 seront également étudiées.

Une analyse complète de l'algorithme de radiosité hiérarchique fera l'objet du chapitre 3. Elle révélera quelques points faibles et donnera deux améliorations de cet algorithme permettant de réduire considérablement les temps de calcul sans diminuer la précision de l'algorithme.

Nous verrons ensuite, au chapitre 4, comment la notion de dérivée s'étend aux fonctions de plusieurs variables, telles que la radiosité. Les opérateurs de dérivation formelle permettent une reformulation de l'expression de la radiosité qui facilite un calcul approché quelle que soit l'expression de la radiosité sur l'émetteur. Nous verrons aussi qu'il est possible de calculer de façon approchée les dérivées successives de la radiosité quelle que soit la radiosité sur l'émetteur.

Cependant, la connaissance de quantités définies ponctuellement (*locales*) ne permet de déduire des informations valables sur toute une surface (*globales*) qu'en présence d'informations supplémentaires sur la fonction étudiée. Les informations nécessaires au contrôle de l'erreur en radiosité seront introduites au chapitre 5.

L'étude de l'erreur dans les algorithmes de radiosité, à la fois *a posteriori* et au cours des calculs sera l'objet de notre chapitre 6.

C’est à l’aide des informations fournies au chapitre 5 et des méthodes définies au chapitre 6 qu’il est possible de construire un critère de raffinement basé sur un contrôle de l’erreur commise sur chaque interaction. Ce critère et son intégration dans un algorithme de radiosité hiérarchique sera l’objet de notre chapitre 7.

Ce critère de raffinement repose sur un calcul précis des dérivées successives de la radiosité. Ce calcul, en raison du caractère plus technique des méthodes utilisées, a été placé en fin d’ouvrage. Il sera l’objet des chapitres 8, 9 et 10. Nous verrons d’abord le cas d’un émetteur sur lequel la radiosité est constante, en l’absence d’obstacles, avec calcul de la dérivée et de la dérivée seconde et implémentation, puis nous verrons la présence d’obstacles, avec calcul et implémentation de la dérivée, et calcul de la dérivée seconde, et enfin l’extension à un émetteur sur lequel la radiosité n’est pas constante, avec cependant un contrôle de l’erreur commise sur l’approximation.

Première partie

Étude de la méthode de radiosité hiérarchique

2.

La méthode de radiosité hiérarchique

> Et il me suffit ici de vous avertir que les rayons, qui tombent sur les corps qui sont colorés et non polis, se réfléchissent ordinairement de tous côtés, encore même qu'ils ne viennent que d'un seul côté (...)
>
> René DESCARTES, *La Dioptrique*, 1637.

AINSI que le faisait déjà remarquer Descartes, les rayons lumineux arrivant sur un objet en repartent habituellement dans toutes les directions. Aussi l'éclairement d'un point d'un objet influence t-il l'éclairement de tous les objets qui sont visibles de ce point.

Pour quantifier précisément cette influence, il est utile d'avoir recours à un bilan énergétique.

2.1 L'équation de l'éclairage global

En théorie, tout point de l'espace peut émettre de l'énergie lumineuse, soit en lui-même – c'est le cas par exemple d'une flamme de bougie – soit sous l'excitation d'un autre émetteur – c'est le cas d'un gaz phosphorescent ou d'une surface réfléchissante.

L'énergie émise en un point est égale à la somme de l'énergie émise par ce point en propre et à l'énergie émise en ce point sous l'excitation de l'énergie reçue.

Si l'on suppose que le milieu n'a pas d'influence sur les transferts d'énergie, et si l'on suppose en outre que les seuls transferts d'énergie en jeu concernent l'énergie lumineuse (en excluant, par exemple les transformations d'énergie lumineuse en énergie cinétique[1] ou *vice-versa*) le problème subit une première simplification : les seuls émetteurs d'énergie sont les surfaces des objets géométriques présents dans la scène.

Pour modéliser l'image perçue par l'œil, il faut connaître l'énergie émise par les surfaces visibles de l'œil. L'énergie émise par ces surfaces dépend évidemment de l'énergie qu'elles ont reçu.

Une quantité plus aisée à manipuler que l'énergie est la *luminance*, définie comme l'énergie émise dans une certaine direction, par unité de temps, par unité de surface perpendiculaire à la direction de propagation, par unité d'angle solide.

1. La transformation d'énergie lumineuse en énergie cinétique peut se produire dans un radiomètre. La transformation inverse peut se produire, par exemple, lorsque les freins d'une voiture sont chauffés à blanc.

Une propriété utile de la luminance est que la luminance quittant un point x en direction d'un point y, $L(x,y)$ est égale à la luminance arrivant au point y venant du point x.

Par ailleurs, la plupart des récepteurs, tels que l'œil humain ou les appareils photographiques sont sensibles à la luminance et non à l'énergie. La connaissance de la luminance en chacun des points des surfaces de la scène visibles de l'œil est donc suffisante pour obtenir une image.

Dans le cas le plus général, l'équilibre énergétique permet d'exprimer la luminance quittant le point x dans la direction (θ_0, φ_0) comme :

$$L(x,\theta_0,\varphi_0) = L_e(x,\theta_0,\varphi_0) + \int_\Omega \rho_{bd}(x,\theta_0,\varphi_0,\theta,\varphi) L_i(x,\theta,\varphi) \cos\theta d\omega \tag{2.1}$$

- $L(x,\theta_0,\varphi_0)$ est la luminance quittant le point x dans la direction définie par l'angle (en coordonnées sphériques) (θ_0,φ_0) ;
- $L_e(x,\theta_0,\varphi_0)$ est la luminance émise par le point x dans la même direction. C'est une propriété de la surface qui porte le point x ;
- Ω est l'ensemble des directions (θ,φ) possibles (la sphère),
- $L_i(x,\theta,\varphi)$ est la luminance qui arrive au point x depuis la direction (θ,φ) ;
- $L_i(x,\theta,\varphi)\cos\theta d\omega$ est le flux incident élémentaire au point x, provenant de la direction (θ,φ). C'est la puissance par unité de surface qui arrive au point x en provenant de l'angle solide élémentaire $d\omega$ centré autour de la direction (θ,φ) ;
- $\rho_{bd}(x,\theta_0,\varphi_0,\theta,\varphi)$ est la fonction de réflectance bi-directionnelle au point x. C'est – par définition – le rapport de la luminance réfléchie dans la direction (θ_0,φ_0) et du flux incident élémentaire dans la direction (θ,φ) ;
- $d\omega$ est l'angle solide élémentaire autour de la direction (θ,φ).

La réflectance bi-directionnelle est une quantité physique d'unité sr^{-1}, qui peut varier de zéro à l'infini. Une quantité plus intuitive est la réflectance directionnelle, un nombre sans dimension compris entre 0 et 1, qui représente la fraction du flux incident suivant une direction qui repart (dans toutes les directions). La réflectance directionnelle est définie par :

$$\rho_d(x,\theta,\varphi) = \int_\Omega \rho_{bd}(x,\theta_0,\varphi_0,\theta,\varphi)\cos\theta_0 d\omega_0 \tag{2.2}$$

L'équation 2.1 est une équation intégrale. La fonction à calculer L apparaît sous le signe d'intégration dans le terme de droite, aussi bien que dans le terme de gauche. Dans les cas les plus généraux, les équations de ce type n'ont pas de solution analytique, et l'on doit se rabattre sur des approximations numériques.

2.2 Résolution formelle

Toutefois, on peut résoudre formellement l'équation 2.1 en introduisant un opérateur $\mathcal{R}$, tel que :

$$(\mathcal{R}L)(x,\theta_0,\varphi_0) = \int_\Omega \rho_{bd}(x,\theta_0,\varphi_0,\theta,\varphi) L_i(x,\theta,\varphi)\cos\theta d\omega \tag{2.3}$$

L'opérateur $\mathcal{R}$ représente l'effet sur L de la réflexion sur toutes les surfaces qui composent la scène.

L'équation 2.1 s'écrit alors sous forme simplifiée :

$$L = L_e + \mathcal{R}L \tag{2.4}$$

Ce qui peut se résoudre, formellement :

$$(I-\mathcal{R})L = L_e$$

$$L = (I-\mathcal{R})^{-1}L_e$$

Et en utilisant une série de Neumann :

$$L = \sum_{n=0}^{+\infty} \mathcal{R}^n L_e \tag{2.5}$$

Cette résolution formelle a un sens physique : l'opérateur $\mathcal{R}$ représente l'effet sur la luminance d'une réflexion sur toutes les surfaces composant la scène.

Ainsi, $\mathcal{R}^0 L_e = L_e$ est la luminance émise en propre. $\mathcal{R}^1 L_e = \mathcal{R}L_e$ est la luminance qui a été réfléchie une fois par ces surfaces, $\mathcal{R}^2 L_e$ la luminance réfléchie deux fois, et ainsi de suite. La luminance émise au total est égale à la luminance émise en propre plus la luminance réfléchie un fois, plus la luminance réfléchie deux fois, et ainsi de suite.

2.3 Première résolution numérique : le lancer de rayons

La formalisation de l'équation 2.1 date de 1986, et a été faite par James Kajiya [25]. Bien avant cette mise en équation du problème global, dès 1980 en fait (voir Whitted [47]), les chercheurs et les industriels utilisaient une méthode de résolution approchée du problème de l'éclairage, le lancer de rayons.

La simplification apportée est simple : on suppose que tous les objets présents dans la scène sont des réflecteurs parfaits au sens de Descartes, c'est-à-dire qu'un rayon lumineux frappant une surface en faisant un angle θ_i avec la normale à cette surface repart en faisant un angle $\theta_r = \theta_i$ (voir figure 2.1).

Le calcul de l'image de synthèse se fait alors en suivant tous les rayons lumineux dans leur trajet de la source lumineuse jusqu'à l'œil. Les résultats peuvent être spectaculaires et utiles. Cependant, dans certaines applications, le lancer de rayons est victime de la supposition de départ : très peu de surfaces sont des réflecteurs spéculaires parfaits au sens de la loi de Descartes. Il peut en résulter des effets spéciaux indésirables. D'autre part, le lancer de rayons est une méthode de résolution ponctuelle, donnant des informations ponctuelles. On n'a pas de contrôle sur l'exactitude des valeurs calculées.

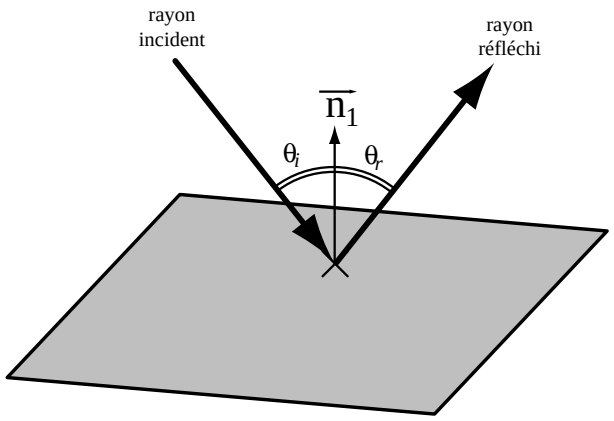


FIG. 2.1 – *Loi de réflexion de Descartes : surfaces spéculaires.*

2.4 Autre simplification : la radiosité

La méthode de radiosité, définie en 1984 par Goral *et al.* [15], introduit une autre simplification pour résoudre l'équation 2.1 : on fait l'hypothèse que toutes les surfaces intervenant dans la scène sont des surfaces diffuses. C'est-à-dire qu'un rayon frappant une surface est réfléchie de façon uniforme dans toutes les directions (voir figure 2.2).

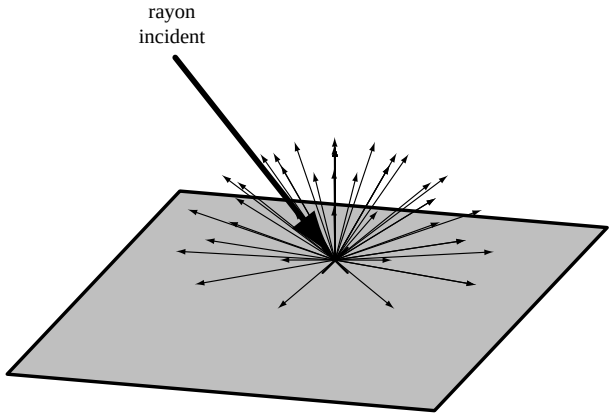


FIG. 2.2 – *Loi de réflexion de Lambert : surfaces diffuses.*

La réflectance bi-directionnelle est par définition indépendante des directions incidentes et réfléchies :

$$\rho_{bd}(x,\theta_0,\varphi_0,\theta,\varphi) \equiv \rho_{bd}(x)$$

La valeur de la réflectance directionnelle que nous avons introduit plus haut (équation 2.2) est ainsi :

$$\rho_d(x) = \rho_{bd}(x) \int_\Omega \cos\theta_0 d\omega_0 = \pi\rho_{bd}(x)$$

Cette réflectance représente la portion de l'énergie arrivant en x suivant une direction donnée qui est renvoyée dans toutes les directions.

Comme la surface est supposée diffuse, la luminance ne dépend également que de la position du point et non des directions :

$$L(x,\theta,\varphi) \equiv L(x)$$

On introduit la radiosité $B(x)$, qui est la puissance en un point, par unité de surface ; c'est-à-dire l'intégrale de la luminance sur toutes les directions :

$$B(x) = \int_\Omega L(x,\theta,\varphi)\cos\theta d\omega$$

La radiosité se simplifie :

$$\begin{aligned} B(x) &= L(x)\int_\Omega \cos\theta d\omega \\ &= L(x)\int_0^\pi \int_0^{2\pi} \cos\theta\sin\theta d\theta d\varphi \\ &= \pi L(x) \end{aligned} \tag{2.6}$$

En reportant $B(x)$ et ρ_d dans l'équation 2.1 :

$$\frac{1}{\pi}B(x) = \frac{1}{\pi}E(x) + \frac{1}{\pi}\rho_d(x)\int_\Omega L_i(x,\theta,\varphi)\cos\theta d\omega \tag{2.7}$$

$E(x)$ est l'exitance, ou la puissance par unité de surface émise au point x. En multipliant l'équation 2.7 par π :

$$B(x) = E(x) + \rho_d(x)\int_\Omega L_i(x,\theta,\varphi)\cos\theta d\omega \tag{2.8}$$

Si on pose :

$$H(x) = \int_\Omega L_i(x,\theta,\varphi)\cos\theta d\omega \tag{2.9}$$

on a :

$$B(x) = E(x) + \rho_d(x)H(x) \tag{2.10}$$

C'est-à-dire que la radiosité émise au point x est égale à la radiosité émise en propre par le point x et au produit de la réflectance au point x par le flux incident au point x venant de toutes les directions possibles.

Cette intégrale sur toutes les directions possibles dépend en fait de la radiosité en tous les points visibles de x. Cette dépendance n'est pas explicite dans l'écriture de $H(x)$ dans l'équation 2.9. Pour l'expliciter, il suffit de transformer l'intégrale sur les directions en une intégrale sur les surfaces.

Si un point y est visible du point x dans la direction (θ,φ), alors x est aussi visible de y, dans une direction (θ',φ'). Or la luminance quittant le point y en direction de x est égale à la luminance arrivant au point x en provenance de la direction de y.

$$L_i(x,\theta,\varphi) = L(y,\theta',\varphi')$$

Comme le point y se trouve sur une surface également diffuse :

$$L(y,\theta',\varphi') = \frac{B(y)}{\pi}$$

L'équation 2.8 peut ainsi être transformée en une intégration sur des surfaces, en explicitant l'angle solide élémentaire :

$$d\omega = \frac{\cos\theta' dy}{r^2}$$

et en prenant pour domaine d'intégration l'ensemble des surfaces qui sont visibles de x. Pour simplifier le domaine d'intégration, il est équivalent d'introduire une fonction de visibilité $V(x,y)$, qui vaut 1 si x et y sont visibles l'un de l'autre, et 0 sinon. Auquel cas,

$$B(x) = E(x) + \rho_d(x) \int_{y\in S} B(y) \frac{\cos\theta\cos\theta'}{\pi r^2} V(x,y)\, dy \tag{2.11}$$

où S est l'ensemble des surfaces composant la scène.

De même que pour $\mathcal{R}$, on peut introduire l'opérateur $\mathcal{H}$ tel que :

$$B(x) = E(x) + \rho_d(x)\, (\mathcal{H}B)\,(x)$$

$$(\mathcal{H}f)(x) = \int_{y\in S} f(y) \frac{\cos\theta\cos\theta'}{\pi r^2} V(x,y)\, dy \tag{2.12}$$

2.5 Discrétisation et première résolution

L'équation 2.11 est impossible à résoudre dans le cas général[2]. La méthode la plus classique de résolution, introduite par Goral *et al.* en 1984 [15] passe par une discrétisation de l'environnement. On décompose la scène S en facettes $(P_i)_{i\in[1,N]}$, sur lesquelles on suppose les paramètres du problème constants.

On note A_i l'aire de la facette P_i, et l'on discrétise une valeur $X(x)$ en prenant sa valeur moyenne X_i sur la facette P_i :

$$X_i = \frac{1}{A_i} \int_{x\in P_i} X(x)$$

L'équation continue 2.11 se discrétise alors en :

$$B_i = E_i + \rho_i \sum_{j=1}^{N} B_j \frac{1}{A_i} \int_{x\in P_i} \int_{y\in P_j} \frac{cos\theta\cos\theta'}{\pi r^2} V(x,y)\, dy \tag{2.13}$$

Si l'on note F_{ij} l'expression de l'intégrale, qui ne dépend que de la géométrie des facettes P_i et P_j, on a :

$$B_i = E_i + \rho_i \sum_{j=1}^{N} F_{ij} B_j \tag{2.14}$$

2. En fait, même des cas très simples, tels que deux objets plans de largeur finie, se touchant suivant un angle droit. Voir, par exemple, Holzschuch, 1994 [22], ou Sillion, 1994 [39].

$$F_{ij} = \frac{1}{A_i} \int_{x \in P_i} \int_{y \in P_j} \frac{cos\theta \cos\theta'}{\pi r^2} V(x,y) dy \tag{2.15}$$

F_{ij}, que l'on appelle le facteur de forme entre les facettes P_i et P_j, représente la proportion de la puissance quittant la facette P_i qui atteint la facette P_j.

Si l'on note $\boldsymbol{B} = [B_i]$ le vecteur formé de toutes les radiosités des facettes créées lors de la discrétisation, et de même $\boldsymbol{E} = [E_i]$ le vecteur des exitances, et si l'on note $\boldsymbol{M} = [F_{ij}]$ la matrice des facteurs de forme, l'équation discrétisée 2.14 s'écrit :

$$\boldsymbol{B} = \boldsymbol{E} + \boldsymbol{M}\boldsymbol{B}$$

On peut résoudre cette équation :

$$(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{B} = \boldsymbol{E}$$

$$\boldsymbol{B} = (\boldsymbol{I} - \boldsymbol{M})^{-1}\boldsymbol{E}$$

$$\boldsymbol{B} = \sum_{i=0}^{+\infty} \boldsymbol{M}^i \boldsymbol{E}$$

La solution du problème discrétisé revient donc à calculer la somme des $\boldsymbol{M}^i\boldsymbol{E}$. Le calcul d'un seul coefficient de $\boldsymbol{M}$ nécessite une estimation de la visibilité entre deux facettes. Cette opération ne peut se faire qu'en temps au moins logarithmique en moyenne, $O(\log m)$, et linéaire dans le cas le pire, $O(m)$, où m est le nombre d'obstacles dans la scène. Le seul calcul de $\boldsymbol{M}$ est donc une étape en $O(N^2 \log m)$ en moyenne, $O(N^2 m)$ dans le cas le pire. Après quoi l'on résout le système en calculant les itérées successives de $\boldsymbol{E}$ par $\boldsymbol{M}$. Chaque itération coûte $O(N^2)$.

Comme dans la résolution théorique (section 2.2), chaque itération représente une réflexion de la lumière sur les surfaces composant la scène : la première itération représente la lumière réfléchie une fois, la deuxième itération la lumière qui a été réfléchie deux fois, et ainsi de suite. La convergence de la méthode de radiosité classique provient du fait que toutes les réflectances sont plus petites que un, car il n'y a pas de création d'énergie au moment de la réflexion, donc que la radiosité diminue à chaque réflexion.

On peut éviter de calculer toute la matrice $\boldsymbol{M}$ lors de la résolution. La plupart des algorithmes de radiosité se contentent d'en calculer une ligne ou une colonne à la fois et de l'utiliser, soit pour propager la radiosité au départ d'une facette, soit pour rassembler la radiosité reçue par une facette (voir figure 2.3).

L'utilisation d'une ligne de la matrice correspond à *rassembler* sur une facette toute la radiosité qu'elle reçoit des autres facettes de la scène, tandis que l'utilisation d'une colonne équivaut à *tirer* la radiosité de la facette correspondante en direction de toutes les autres facettes de la scène. Si l'on trie les facettes en fonction de leur radiosité et que l'on tire les plus brillantes d'abord, l'algorithme converge bien plus rapidement (voir Cohen, 1988 [6]).

2.6 Radiosité hiérarchique

2.6.1 Principes

L'un des problèmes de la discrétisation décrite dans 2.5 est que l'on a décomposé chaque objet à un seul niveau de détail, indépendamment de sa position par rapport à d'autres objets.

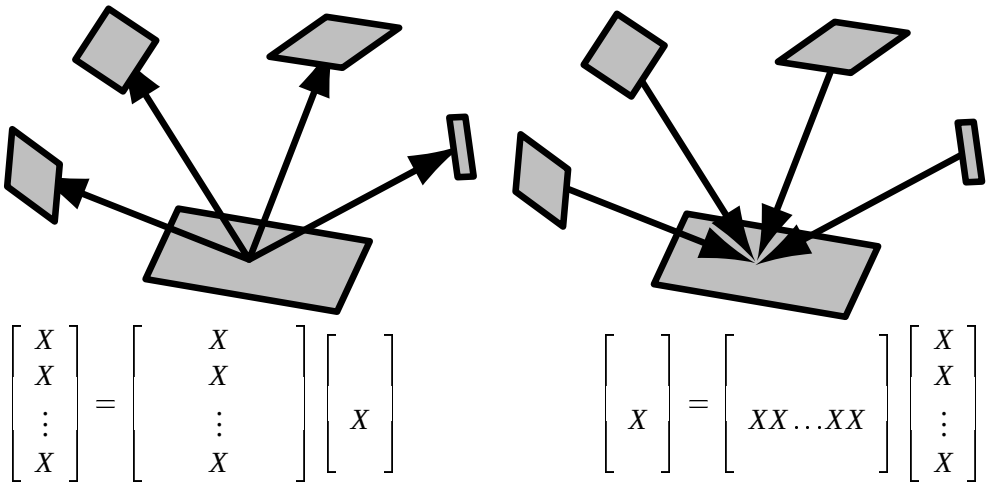


FIG. 2.3 – *Utilisation d'une ligne ou d'une colonne de* ***M****.*

Supposons que nous nous intéressions à la valeur de la radiosité en un point *x*. Nous avons vu plus haut que cette radiosité était l'intégrale, pour toutes les surfaces visibles de ce point, de leur luminance, multipliée par l'angle solide sous lequel elles sont visibles de *x*. Ce qui nous intéresse est une estimation de cette intégrale, basée sur une discrétisation des surfaces visibles. On comprend que plus une surface est proche du point *x*, plus il sera utile de l'avoir discrétisée finement. Inversement, plus un objet est éloigné et plus on peut se permettre une discrétisation grossière. Si le niveau de discrétisation d'une surface est indépendant de l'objet avec lequel elle interagit, on court le risque, soit de manquer de précision en certains points, soit d'effectuer des calculs inutiles en d'autres.

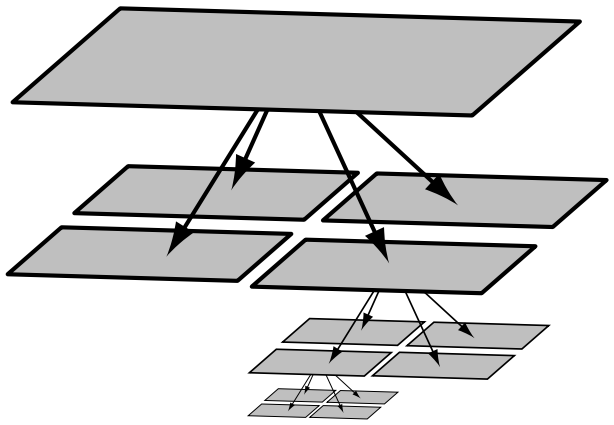


FIG. 2.4 – *Radiosité hiérarchique : chaque objet est décomposé en une hiérarchie de facettes.*

La méthode de radiosité hiérarchique, introduite par Patrick Hanrahan en 1990 [16], discrétise les objets de façon hiérarchique. Chaque surface géométrique de la scène donnée en entrée est décomposée en une hiérarchie de facettes (voir figure 2.4).

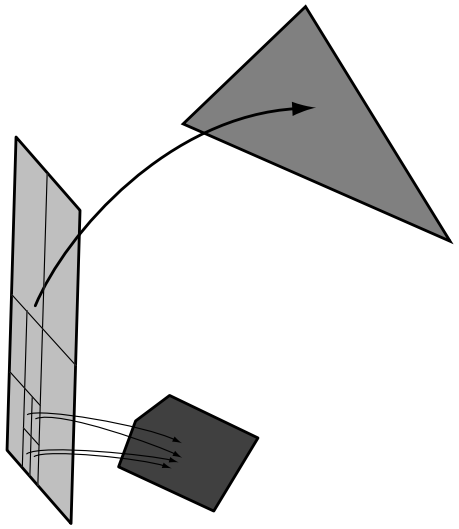


FIG. 2.5 – *Radiosité hiérarchique : les interactions entre les objets.*

L'interaction entre deux surfaces est modélisée par des liens entre les différentes facettes qui composent ces surfaces (voir figure 2.5). Ainsi, l'interaction entre deux facettes est-elle modélisable à différents niveaux de précision.

Comme la discrétisation du problème et la méthode de résolution sont aussi des sources d'erreurs, il n'est pas nécessaire de modéliser l'interaction entre objets *au-delà* de la précision minimale des autres parties de l'algorithme.

2.6.2 Structures de données

Au départ, l'algorithme de radiosité hiérarchique ne connaît de la scène que sa définition géométrique. Ainsi, un plancher de 5 mètres sur 5 est initialement donné comme un seul objet. Au cours de la résolution itérative, chaque objet géométrique porte une hiérarchie de facettes, chaque facette ayant une radiosité distincte. Cette hiérarchie est organisée sous la forme d'un arbre : la racine de l'arbre porte l'objet géométrique de base ; chacun des nœuds de l'arbre porte à la fois une partie de cet objet, la radiosité estimée pour cette partie, et la liste des interactions de cette facette avec le reste de la scène.

Les interactions sont stockées sous forme de liens : chaque lien relie deux facettes, une facette source et une facette destination, et contient une estimation du facteur de forme de la facette destination vers la facette source. En vertu du principe du retour inverse de la lumière, s'il existe un lien de la facette p vers la facette q, il existe nécessairement un autre lien de la facette q vers la facette p.

2.6.3 Algorithme

L'algorithme général est simple :

1. lors d'une étape de pré-traitement, on examine la situation de visibilité de tous les couples d'objets géométriques, et on établit un lien entre les racines des arbres de ces objets dès qu'ils ne sont pas mutuellement invisibles ;

2. à chaque étape du calcul, on examine tous les liens de la scène, et l'on raffine ceux d'entre eux qui ne satisfont pas un certain critère de précision ;

3. puis l'on propage l'énergie le long des liens établis ;

4. enfin, pour chaque arbre, on propage l'énergie reçue aux différents niveaux dans la hiérarchie afin d'uniformiser la représentation.

La dernière étape nécessite une explication : si chaque nœud de l'arbre porte une estimation de la radiosité moyenne pour sa facette, des interactions portées par les nœuds ascendants ou descendants influencent également la radiosité du nœud. C'est cette influence qu'il faut estimer pour que la radiosité portée par le nœud soit vraiment une moyenne. Toute la radiosité reçue par le nœud « père » lors de l'étape de propagation est passée à ses « fils », tandis que le « père » reçoit la moyenne de la radiosité reçue par ses « fils ».

Cette propagation de l'information se fait par un parcours en profondeur d'abord de chacun des arbres concernés.

Par ailleurs, la propagation de l'énergie dans la scène (étapes 3 et 4) le long des liens correspond à une multiplication du vecteur des radiosités $\boldsymbol{B}$ par la matrice $\boldsymbol{M}$, qui est ajoutée à la valeur ancienne de $\boldsymbol{B}$. La valeur de $\boldsymbol{B}$ à l'étape k est donc :

$$\boldsymbol{B}_k = \sum_{i=0}^{k} \boldsymbol{M}^i \boldsymbol{B}_0$$

Si $\boldsymbol{B}_0$ est le vecteur des émittances $\boldsymbol{E}$, on voit que la méthode de radiosité hiérarchique converge vers la solution :

$$\boldsymbol{B} = (\boldsymbol{I} - \boldsymbol{M})^{-1} \boldsymbol{E} = \sum_{i=0}^{+\infty} \boldsymbol{M}^i \boldsymbol{E}$$

2.6.4 Oracles

L'un des avantages de l'algorithme de radiosité hiérarchique est qu'il raffine les objets au fur et à mesure de la résolution, en fonction de la précision nécessaire. Toute la difficulté est dans l'estimation de la précision actuelle de la modélisation, afin de prendre la décision de raffiner ou non l'interaction.

– Le premier algorithme, proposé par Hanrahan en 1990 [16], partait de la constatation que les différentes méthodes d'estimation du facteur de forme étaient précises dans les cas où les objets étaient de taille petite devant leur distance. Autrement dit, dans les cas où l'angle solide sous-tendant la surface était petit ou dans les cas où le facteur de forme lui-même était petit. L'oracle de raffinement était alors un test sur l'estimation du facteur de forme par rapport à un certain ε_F ;

– Le problème de cet oracle est qu'on peut raffiner fortement une interaction qui n'aura pas d'influence sur l'éclairage de la scène, par exemple une interaction entre deux surfaces peu éclairées. D'où l'idée d'un deuxième oracle, proposé par Hanrahan en 1991 [17] : le raffinement est basé sur la comparaison du produit de la radiosité et du facteur de forme avec ε_{BF}. C'est-à-dire que l'on raffine une interaction seulement si l'imprécision sur la radiosité qu'elle transporte dépasse un certain seuil ;

– Un troisième oracle de raffinement, introduit par Smits en 1992 [40], utilise le concept d'importance : l'importance $\boldsymbol{Z}$ est une quantité duale de la radiosité, qui est solution d'une équation de transport duale de celle de la radiosité.

$$\begin{aligned} \boldsymbol{B} &= \boldsymbol{E} + \boldsymbol{M}\boldsymbol{B} \\ \boldsymbol{Z} &= \boldsymbol{R} + {}^{t}\boldsymbol{M}\boldsymbol{Z} \end{aligned}$$

$\boldsymbol{R}$ est le vecteur des émissions d'importance, dit aussi vecteur de réception. Ainsi, l'importance Z_p d'une facette p correspond-elle à l'influence qu'aura cette facette sur les facettes qui émettent de l'importance (celles telles que $R_i \neq 0$). Un cas particulier très simple et très efficace consiste à prendre pour unique facette émettrice d'importance une facette placée à l'emplacement de l'œil ou de la caméra qui observe la scène. Naturellement, l'importance n'est pas limitée à ce seul cas particulier (voir section 6.8.2).

2.6.5 Complexité

La complexité de l'algorithme de radiosité hiérarchique est un point important, et c'est là-dessus que repose sa supériorité sur les autres algorithmes.

– Étant donné une scène de m objets, le pré-traitement nécessite l'établissement de $\frac{m(m-1)}{2}$ relations. Chacune de ces relations nécessite un calcul de visibilité, donc on a une complexité de l'étape préliminaire en $O(m^2 \log m)$, voire $O(m^3)$ en temps et $O(m^2)$ en mémoire ;

– Lorsqu'on raffine une interaction, on remplace l'ancienne interaction par quatre interactions, d'une part, et on ajoute quatre nouvelles facettes d'autre part. Le nombre d'interactions est donc lié au nombre de facettes terminales, soit de feuilles dans la hiérarchie, lequel est lié au nombre total de facettes, c'est-à-dire de nœuds dans la hiérarchie. Si N est le nombre total de facettes, le nombre d'interactions est donc $O(N)$. La complexité d'une étape de raffinement et propagation est ainsi $O(N)$. Ce résultat est à comparer avec la méthode de radiosité classique, où la matrice des facteurs de forme contient N^2 coefficients. On a ici modélisé une matrice N^2 par $O(N)$ blocs ;

– La complexité globale de l'algorithme dépend donc à la fois de m et de N. Chaque étape de raffinement peut nécessiter un calcul de visibilité – lequel est en $O(\log m)$, voire $O(m)$. La place mémoire est en $O(m^2 + N)$. La relation entre m et N dépend de la géométrie de la scène concernée. Si la scène comprend de nombreuses petites surfaces visibles les unes des autres, $N \approx m$, et la radiosité hiérarchique n'a rien apporté. Si la scène comprend de grands polygones qu'il faut raffiner, $N \gg m$.

Ainsi, dans le pire des cas, la méthode de radiosité hiérarchique est en $O(m^3 + Nm)$, alors que la méthode classique est en $O(N^2 m)$.

3.

Améliorations de l'algorithme original de radiosité hiérarchique, basées sur l'analyse

NOUS présentons ici une analyse détaillée[1] de l'algorithme de radiosité hiérarchique. Elle confirme que les tests de visibilité sont la partie la plus coûteuse de l'algorithme. Par ailleurs analyser le comportement de l'algorithme permet de construire deux améliorations sensibles. Une évaluation «paresseuse» des liens entre les surfaces de haut niveau définissant la scène élimine la plus grande partie du travail de pré-traitement nécessaire à l'évaluation de ces liens. Qui plus est, le nombre de liens nécessaires pour modéliser l'interaction entre deux surfaces mutuellement visibles peut être réduit grâce à un nouvel oracle de raffinement. Les résultats expérimentaux montrent que le travail d'établissement des liens initiaux peut être largement évité et le nombre de liens substantiellement réduit sans perte de qualité pour l'image produite, et que ces modifications accélèrent sensiblement l'algorithme de radiosité hiérarchique.

3.1 Motivations

Pour étudier le comportement de l'algorithme de radiosité hiérarchique, nous avons lancé le programme de radiosité hiérarchique original (décrit par Hanrahan [16, 17]) sur un ensemble de cinq scènes d'intérieur, de complexité croissante (de 170 à 2355 polygones initiaux). Ces scènes proviennent de différents programmes de recherche, et ont des aspects géométriques notablement différents. En les utilisant, nous voulons identifier des propriétés générales des scènes d'intérieur, et éviter ainsi le problème classique d'une généralisation abusive basée sur l'étude d'une seule scène ou d'un seul ensemble de scènes similaires.

Les scènes de test sont «Bureau», qui est la scène originale utilisée dans l'article de Hanrahan [17], «Salle à manger» qui est la scène n° 7 d'un ensemble de 10 scènes de référence distribuées par Peter Shirley pour le *Fifth Eurographics Workshop on Rendering*, «Chercheur» et «Stagiaire» qui sont des scènes contenant des modèles de bureaux et de chaises légèrement plus sophistiqués, et finalement «Hévéa», qui contient un modèle complexe d'arbre, et de nombreux problèmes de visibilité mutuelle. Le tableau 3.1

1. Une partie de ce travail a fait l'objet d'une publication en anglais au *Fifth Eurographics Workshop on Rendering*, à Darmstadt en juin 1994 [23].

TAB. 3.1 – *Description des cinq scènes de test.*

Nom	*n*	Description
Bureau	170	Un bureau
Salle à manger	402	Une table et quatre chaises
Chercheur	1006	Deux bureaux et six chaises
Stagiaire	1647	Quatre bureaux et dix chaises
Hévéa	2355	Un hévéa et trois sources lumineuses

donne pour chaque scène le nombre *n* de polygones initiaux et une brève description. Voyez les figures 3.9, 3.13 et 3.14 pour un aperçu de ces scènes.

Les paramètres du programme ont, pour chaque scène, été ajustés afin de donner la meilleure image possible le plus rapidement possible. Cet ajustement fait, en un sens, que les conditions expérimentales n'étaient pas les mêmes pour les différentes scènes. D'un autre côté, les échelles des scènes étant différentes à cause de leurs origines variées, les paramètres basés sur ces dimensions, au moins, devaient être modifiés. Dans l'idéal, il aurait fallu pouvoir lancer le programme avec les mêmes paramètres pour chaque scène.

3.2 La visibilité

La première observation tirée de l'analyse de l'algorithme de radiosité hiérarchique est la quantification de l'importance des calculs de visibilité dans le coût total de l'algorithme de radiosité hiérarchique. Ainsi qu'il l'a souvent été postulé dans les travaux antérieurs, tels que ceux de Teller et Lischinski en 1993 [41, 42, 27], les calculs de visibilité représentent une portion très importante du coût total. La figure 3.1 montre le temps passé par le programme dans les différentes étapes de l'algorithme de radiosité hiérarchique.

Pour présenter les résultats de façon lisible, l'algorithme a été découpé en étapes correspondant aux étapes logiques de l'algorithme. Ainsi, «Visibilité» correspond aux calculs de visibilité effectués, «Facteur de Forme » aux calculs de facteurs de forme en dehors de la visibilité[2]. «Raffinement » est le temps passé par l'algorithme dans la procédure de raffinement des liens (y compris les différents oracles et la subdivision des facettes) ; «Propagation» correspond au temps mis pour propager l'énergie le long des liens de facette à facette établis au cours de l'étape de raffinement ; enfin, «Hiérarchie» correspond à la propagation de l'énergie reçue dans la hiérarchie de facettes construite sur chaque polygone, ce qui se fait par une parcours en profondeur d'abord de chaque arbre.

Cette dernière étape est très simplement proportionnelle au nombre total de facettes créées, indépendamment de la position des différents objets dans la scène ou de la quantité d'énergie échangée.

Il est important de noter que très rapidement (au bout de quatre à six itérations) l'al-

2. L'algorithme utilisait pour cela une approximation point-disque.

gorithme a atteint un point ou l'effet des itérations successives n'est plus visible à l'œil nu. Dans de telles itérations, on ne raffine plus, et donc on ne fait plus de tests de visibilité. En revanche, on continue à propager l'énergie le long des liens créés et dans la hiérarchie de chaque polygone. L'effet de ces itérations est donc d'augmenter l'importance relative des étapes «Hiérarchie» et «Propagation», au détriment de «Visibilité», «Raffinement» et «Facteur de Forme». Cette déformation est toutefois très relative car ces étapes sont très courtes si on les compare aux premières étapes.

Notons ici qu'il est très difficile de tester automatiquement la convergence d'un algorithme de synthèse d'image : on peut éventuellement tester la différence entre l'étape n et l'étape $n+1$, mais il peut se produire que cette différence soit infiniment petite sans pour autant que l'on ait atteint la convergence. C'est pour cette raison que, pour les tests, notre algorithme effectuait une nombre fixé à l'avance d'itérations (ici, dix).

Nous avons vu que pour nos scènes de test, au bout de quatre à six itérations, on ne distinguait plus de différences à l'œil nu. Si donc l'algorithme de radiosité hiérarchique est utilisé au sein d'un programme interactif où l'utilisateur fixe lui-même s'il veut continuer les calculs ou non, l'importance relative de la visibilité, du raffinement et du calcul de facteur de forme augmenterait encore.

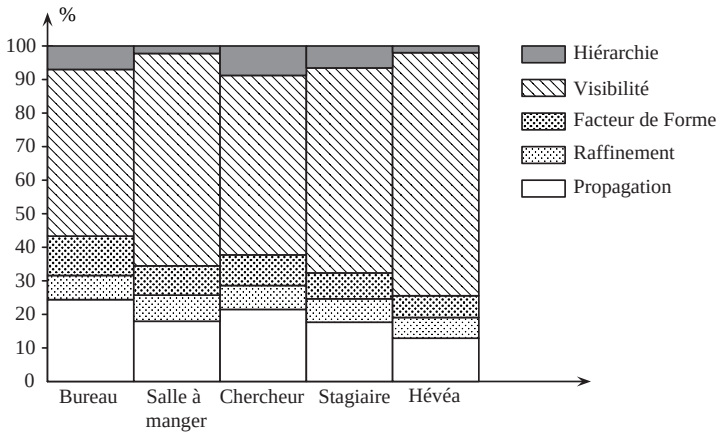


FIG. 3.1 – *Les temps relatifs de chaque étape.*

Le graphe de la figure 3.1 montre que la visibilité domine les calculs en radiosité hiérarchique. Naturellement, ceci dépend de l'algorithme de visibilité utilisé. L'importance de la visibilité pourrait être réduite en utilisant un pré-traitement de la scène, comme dans l'algorithme décrit par Teller en 1993 [42] ou l'algorithme de visibilité multi-résolution décrit par Sillion *et al.* en 1995 [38]. Cependant, l'importance quantitative de la visibilité sur les autres étapes de l'algorithme tient aussi à la nature de l'algorithme.

En effet, à chaque itération, pour un lien donné, on appelle la procédure de raffinement. La décision de raffiner ce lien est prise en fonction de données déjà calculées et surtout qui ne dépendent que des deux polygones reliés. Quand la décision de raffiner est prise, cela déclenche un calcul de facteur de forme – qui, là encore, ne dépend que des deux polygones reliés – et un calcul de visibilité, qui, lui, dépend *potentiellement*

de tous les polygones présents dans la scène, même si en général les techniques d'accélération (grilles, arbres BSP) permettent de simplifier les calculs. On a coutume de dire qu'un calcul de visibilité dépend en moyenne du logarithme du nombre de polygones présents dans la scène, même si pour toute méthode d'accélération il est possible de construire une scène ou un calcul de visibilité dépendra de tous les polygones.

Pour chacun des liens créés au cours du raffinement progressif, on a eu un appel de « Raffinement », un appel de « Facteur de Forme » et un appel de « Visibilité ». De même, lorsqu'on propage l'énergie, « Propagation » sera appelée une fois pour chacun des liens créés. Mais «Propagation» n'utilise que le lien, le facteur de forme qu'il porte et les deux polygones qu'il relie. Enfin, on a vu que le nombre d'appels à «Hiérarchie» ne dépend que du nombre de facettes créées, et que le nombre de facettes créées dépend linéairement du nombre de liens créés.

Ces cinq procédures sont donc appelées un nombre de fois sensiblement équivalent au cours de la résolution, $O(N)$, si N est le nombre de liens créés. Mais de ces cinq procédures, quatre se font en temps constant, et une seule, « Visibilité », a une complexité approximativement logarithmique. Ce qui explique le rôle prépondérant pris par la visibilité dans le temps de calcul, et d'autant plus important que la complexité de la scène augmente.

In fine, cette analyse de l'algorithme de radiosité hiérarchique montre les points où une économie de temps est le plus susceptible d'accélérer les calculs : divisez par deux le temps passé à calculer la visibilité, et vous aurez gagné un tiers du temps total. Inversement, employer une méthode précise de calcul du facteur de forme, même si elle est deux fois plus lente, n'entraîne qu'une perte de temps marginale. Cette perte de temps étant compensée par le fait qu'il est moins nécessaire de raffiner pour obtenir la même précision dans la résolution.

3.3 Bilan énergétique

3.3.1 Principes des calculs

Un des principaux atouts de la méthode de radiosité est sa capacité à modéliser les échanges d'énergie. En effet, toute l'énergie lumineuse quittant un point est prise en compte par l'algorithme, et est transmise à un autre point de la scène. Dans une scène fermée, il n'y a ni création, ni déperdition d'énergie dans la résolution. C'est ce qui rend la méthode de radiosité intéressante pour des usages industriels.

Cette conservation de l'énergie au niveau global correspond à une propriété simple au niveau local : comme le facteur de forme F_{pq} mesure la proportion de l'énergie quittant p qui arrive sur q, en théorie, la somme des facteurs de forme partant d'une facette donnée doit être égale à un. Dans la méthode de radiosité classique, cela signifie que la somme des termes d'une ligne de la matrice des facteurs de forme est égale à un :

$$\forall i,\ \sum_{i=1}^{N} F_{ij} = 1$$

Dans la méthode de radiosité hiérarchique, le calcul est plus compliqué puisqu'une facette donnée émet de l'énergie à la fois en propre et comme membre d'une hiérarchie.

L'énergie quittant une facette dans se décompose en

a) l'énergie qui quitte la facette en propre, en suivant les liens qui partent de la facette même ;

b) l'énergie qui quitte les parents de la facette dans la hiérarchie, modulée en fonction du rapport des surfaces ;

c) l'énergie qui quitte les enfants de la facette dans la hiérarchie.

Naturellement, l'énergie qui quitte les parents de la facette ne peut pas être prise en compte telle quelle comme énergie quittant la facette. En supposant la répartition de l'énergie uniforme sur la facette parente, on doit tenir compte du rapport des surfaces des deux facettes. Ainsi, la somme des facteurs de forme quittant une facette en radiosité hiérarchique peut-elle être calculée comme

a) la somme des facteurs de forme des liens quittant la facette ;

b) plus la somme des facteurs de forme des liens quittant les descendants de la facette ;

c) plus la somme des facteurs de forme des liens quittant les parents de la facette, pondérée par le rapport des surfaces.

Le calcul de la somme des facteurs de forme pour chaque facette peut être fait avec un parcours en profondeur d'abord de l'arbre.

Une fois connue la somme des facteurs de forme pour chaque facette, l'énergie de la facette étant égale au produit de la radiosité par la surface, on peut évaluer l'erreur commise sur l'énergie qui quitte la facette :

$$\mathcal{E}_p = |1 - F_p| \, B_p A_p \tag{3.1}$$

où A_p est la surface du polygone p, B_p sa radiosité et F_p la somme des facteurs de forme quittant p : $F_p = \sum_q F_{pq}$.

Cette quantité exprime l'écart à l'équilibre du bilan énergétique de la facette p. C'est, en quelque sorte, la quantité d'énergie créée ou perdue à chaque itération à cause des erreurs de précision dans le calcul des facteurs de forme. Cette quantité est bien plus significative que la simple différence $|1 - F_p|$, puisque des facteurs de forme très imprécis mais qui ne transportent pas d'énergie ont moins d'impact sur l'image finale et sur le bilan énergétique que ceux qui quittent une facette très éclairée et très énergétique.

On peut aussi calculer l'erreur commise au total pour l'ensemble de la scène :

$$\mathcal{E}_{E_T} = \sum_p |1 - F_p| \, B_p A_p \tag{3.2}$$

que l'on peut comparer avec l'énergie totale de la scène :

$$E_T = \sum_p B_p A_p \tag{3.3}$$

La figure 3.2 montre l'évolution du rapport $\mathcal{E}_{E_T} / E_T$ pour les scènes «Salle à manger» et «Bureau» au cours des calculs.

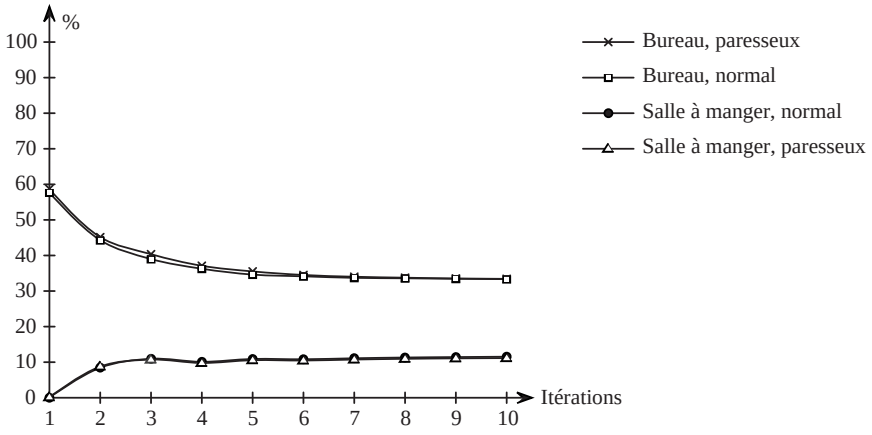


FIG. 3.2 – *Les pertes relatives d'énergie $\mathcal{E}_{E_T}/E_T$.*

3.4 L'étape initiale

Chaque étape de l'algorithme de radiosité hiérarchique consiste à :

– raffiner les liens déjà créés ;

– propager l'énergie le long des liens existants, y compris ceux qui viennent d'être créés ;

– propager l'énergie reçue dans la hiérarchie de chaque polygone.

Ces étapes reposent sur une étape initiale, qui crée les premiers liens entre tous les polygones qui sont au moins partiellement visibles les uns des autres. Cette étape initiale requiert $O(m^2)$ tests de visibilité ; son coût est donc approximativement $m^2 \log m$.

TAB. 3.2 – *L'importance de l'étape initiale dans le temps de calcul total (en secondes).*

Nom	n	Temps total	Étape initiale	Rapport (en %)
Bureau	170	301	5.13	1,7 %
Salle à manger	402	4824	436	9 %
Chercheur	1006	587	194	33 %
Stagiaire	1647	1017	476	47 %
Hévéa	2355	4253	1597	37 %

Le tableau 3.2 présente les temps de calculs pour les scènes de test. Comme on l'a vu plus haut, le temps de calcul total correspond à dix étapes de propagation.

On voit d'après ces statistiques que le coût de l'étape initiale augmente de façon significative avec le nombre d'objets présents dans la scène. L'importance de la structure de la scène est également clairement visible, puisque «Salle à manger» a un temps

de calcul équivalent à « Stagiaire », tout en ayant moins de polygones. Pour toutes les scènes de plus de 1000 polygones, il est clair que le coût de l'étape initiale est important. Ce coût est particulièrement indésirable du fait qu'il est concentré au début du calcul, et qu'ainsi il retarde le moment où une première image utilisable de la scène sera visible.

On note d'ailleurs que les améliorations de l'algorithme de radiosité hiérarchique, telles que les méthodes de Smits *et al.* en 1992 [40] ou Lischinski *et al.* en 1993 [27], tout en réduisant de façon significative le temps de raffinement, continuent à reposer sur la même étape initiale ; cette étape finit par devenir le point le plus coûteux de l'algorithme. Ainsi, les résultats de Smits [40] font état d'un temps total de raffinement de 16 minutes, à comparer avec une étape initiale de 2 heures et 16 minutes.

Une autre observation intéressante peut être faite concernant les liens créés au cours de cette étape initiale, et qui relient les polygones initialement présents dans la scène, avant tout raffinement. C'est le nombre de ces liens pour lesquels le produit BF ne dépasse jamais le seuil de raffinement au cours du calcul. La figure 3.3 montre pour chaque scène la proportion des liens créés au cours de l'étape initiale n'ayant pas été raffinés. On voit qu'un grand nombre de ces liens ne dépassent jamais les critères de raffinement : après 10 itérations, entre 65 % et 95 % d'entre eux n'ont pas été raffinés. Bon nombre de ces liens n'ont sans doute pas ou peu d'impact sur la solution de radiosité calculée.

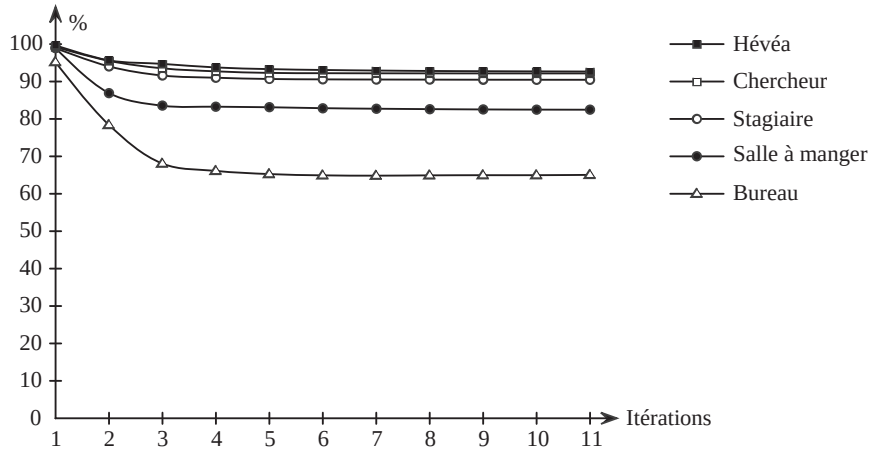


FIG. 3.3 – *Proportion des liens pour lesquels BF ne dépasse pas* ε.

Que tirer des observations précédentes ? D'abord que si l'on pouvait éviter les calculs de l'étape initiale, les premières images utilisables produites par l'algorithme apparaîtraient plus vite. Ensuite que si l'on ne calculait les liens entre les polygones initiaux que lorsqu'on est sûr qu'ils seront utiles, on pourrait économiser une large part des tests de visibilité initiaux.

3.5 Évaluation paresseuse des interactions entre polygones de haut niveau

3.5.1 Principe

Nous proposons ici une modification de l'algorithme de radiosité hiérarchique, qui retarde la création des liens entre polygones de haut niveau jusqu'à ce qu'un tel lien soit supposé nécessaire. L'idée générale est d'éviter d'effectuer des calculs qui n'ont pas d'impact appréciable sur la solution générale, pour se concentrer sur les transferts d'énergie les plus importants.

Cette modification nous impose d'introduire un nouveau critère, qui décidera si deux polygones de haut niveau doivent être liés et interagir ou non. Au cours des calculs, les liens entre polygones de haut niveau sont créés de façon *paresseuse*, uniquement lorsqu'ils semblent nécessaires. Le critère simple employé dans notre implémentation, Holzschuch *et al.* [23] était une adaptation du critère de raffinement des liens entre polygones, donc un test sur le produit BF, avec un seuil propre, $\varepsilon_{\text{link}}$.

3.5.2 Description de l'algorithme

Dans l'algorithme de radiosité hiérarchique, deux polygones sont soit mutuellement invisibles – et donc n'interagissent pas – soit au moins en partie visibles l'un de l'autre, et échangent donc de l'énergie. Nous introduisons une deuxième classification : une paire de polygones est soit *connue*, soit *inconnue*, suivant qu'on possède sur elle des informations ou non. Les paires connues sont celles pour lesquelles on connaît la nature de l'interaction. Initialement, toutes les paires sont donc considérées comme inconnues.

À chaque itération, toutes les paires inconnues sont examinées : on compare d'abord la radiosité des deux polygones avec un seuil, $\varepsilon_{\text{link}}$. Si les polygones sont suffisamment éclairés, on teste s'ils se font face (ce qui peut se faire en temps constant). Si non, la paire est considérée comme connue, et on ne crée pas de liens. Si oui, on calcule une approximation du facteur de forme (sans tests de visibilité) et on compare le produit de la radiosité et du facteur de forme avec $\varepsilon_{\text{link}}$. Si le test est concluant, on marque la paire comme connue et on calcule la visibilité mutuelle des deux polygones. Suivant le résultat du test de visibilité, on crée ensuite un lien reliant les deux polygones.

On voit que les tests de visibilité n'ont lieu qu'une fois que la décision de classer la paire comme connue a été prise. Une paire de polygones n'est connue que lorsque qu'un lien a été établi, ou lorsqu'il est démontré que les deux polygones sont mutuellement invisibles. La figure 3.4 montre une version en pseudo-code de l'étape initiale et d'une itération de l'algorithme original [17], et la figure 3.5 l'équivalent pour l'algorithme proposé ici.

Stocker le statut connue/inconnue de toutes les paires de polygones nécessite $\frac{m(m-1)}{2}$ bits de mémoire. Ce coût représente 62 ko pour une scène d'un millier de surfaces initiales, 25 Mo pour vingt mille surfaces. Ce coût peut devenir prohibitif pour des scènes réellement complexes ; il est cependant comparable au coût global de l'algorithme, $O(m^3 + Nm)$, et pour des scènes de plusieurs dizaines de milliers de polygones, c'est sans doute l'ensemble de l'algorithme de visibilité qu'il faudrait revoir, par exemple en introduisant des méthodes de type *clustering*, regroupant les objets avant de les lier les uns aux autres (voir Sillion, 1994 [36, 37]).

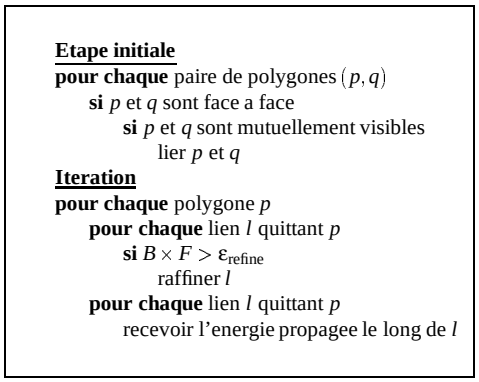


FIG. 3.4 – *Algorithme original.*

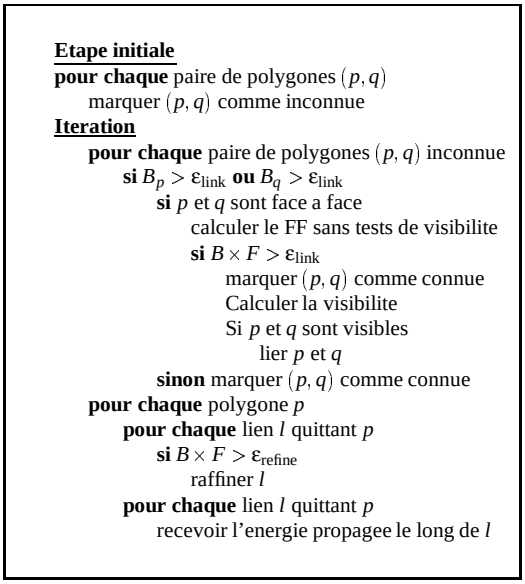


FIG. 3.5 – *Pseudo-code pour notre algorithme.*

Le seuil $\varepsilon_{\text{link}}$ employé dans le critère de création des liens n'est pas nécessairement le même que le seuil $\varepsilon_{\text{refine}}$ employé pour raffiner les liens déjà créés. La principale source d'erreur de notre algorithme est une balance énergétique faussée. Comme l'énergie est transférée le long des liens, un polygone qui n'a pas été lié à tout l'environnement conserve par devers lui de l'énergie, qui n'est pas propagée dans la scène.

À la fin du processus de raffinement des liens, lorsque le produit *BF* devient inférieur à $\varepsilon_{\text{refine}}$, on crée un lien qui sera utilisé pour propager l'énergie du polygone. En revanche, quand deux polygones ne sont pas liés parce que leur produit *BF* est inférieur à $\varepsilon_{\text{link}}$, toute l'énergie qui pourrait être propagée suivant le lien qui les relierait est perdue. Il est donc nécessaire de choisir $\varepsilon_{\text{link}} < \varepsilon_{\text{refine}}$.

Par ailleurs, le simple fait de ne pas créer certains liens et donc de ne pas calculer certains facteurs de forme compromet naturellement le bilan énergétique de la scène. Pour mesurer l'intérêt réel de la modification proposée, nous devons à la fois calculer le temps qu'elle permet de gagner, et la perte de précision qu'elle introduit sur le plan énergétique. Cette perte d'énergie doit notamment être comparée avec la perte de l'algorithme avant la modification : les deux quantités sont-elles comparables – auquel cas la modification est légitime – ou y a-t-il une perte supplémentaire de précision appréciable introduite par notre modification ?

Le résultat dépend naturellement des paramètres du problème, et en particulier de la fonction utilisée pour approximer le facteur de forme. Avec une approximation point-disque, nous trouvons que la modification de l'algorithme n'introduit pas de pertes de précision supplémentaires (voir figure 3.2).

Si la perte de précision se trouve être appréciable (dans des conditions à établir), la solution que je propose est de rétablir le bilan énergétique en introduisant un terme ambiant similaire à celui proposé par Cohen en 1988 [6], calculé à partir de l'ensemble de l'énergie excédentaire dans la scène, et qui serait redistribuée aux différents objets.

Nous avons vu que la somme des facteurs de forme portés par les liens qui quittent un objet est théoriquement égale à un. Si, dans la pratique cette somme est différente de un, c'est parce que les liens qui n'ont pas été créés à cause de notre évaluation paresseuse font défaut. On peut considérer que l'objet considéré a une partie de son énergie qu'il ne peut diffuser, et qu'inversement il y a une partie de l'énergie présente dans la scène qu'il ne peut recevoir.

On commence par calculer l'énergie à diffuser dans la scène : c'est la somme, pour toutes les facettes, de l'énergie qu'elles n'ont pas pu diffuser, calculée en prenant la différence avec un de la somme des facteurs de forme, et en multipliant par la radiosité de la facette et par son aire. L'énergie non diffusée pour une facette est :

$$\mathcal{E}_p = (1 - \sum_q F_{pq}) B_p A_p$$

et pour l'ensemble de la scène, l'énergie qui reste à diffuser est :

$$\mathcal{E} = \sum_p F_p$$

L'énergie à diffuser est divisée par la somme des aires des objets présents dans la scène, ce qui nous donne la radiosité de notre émetteur fictif :

$$\mathcal{B} = \frac{\mathcal{E}}{\sum_p A_p}$$

La redistribution est aussi basée sur la somme des facteurs de forme, contrairement à ce qui est proposé dans Cohen, 1988 [6]. En effet, si la somme des facteurs de forme d'un objet est égale à un, ses échanges avec le monde extérieur sont correctement modélisés, et nous ne voulons pas modifier sa radiosité avec notre terme ambiant.

Chaque objet reçoit donc le terme ambiant, $\mathcal{B}$, multiplié par la différence entre 1 et la somme de ses facteurs de forme :

$$dB = \mathcal{B}(1 - \sum_q F_{pq})$$

Ce qui revient à dire que l'on crée un émetteur fictif, de radiosité $\mathcal{B}$, et tel que le facteur de forme entre lui et une facette p soit égal à $(1 - \sum_q F_{pq})$. De la sorte, la somme des facteurs de forme pour chaque facette est bien égal à un.

Ce terme ambiant, comme celui de Cohen, 1988, n'est utilisé que lors de l'affichage de la scène, après la fin des calculs d'une itération. Il n'est pas utilisé dans les transferts d'énergie entre facettes.

3.6 Raffinement excessif

La quatrième observation importante tirée de notre analyse de l'algorithme de radiosité hiérarchique est la subdivision excessive qui se produit, particulièrement dans les zones totalement éclairées. Ce dernier point est plus délicat à quantifier que les deux autres. Pour l'illustrer, prenons une vue de la scène « Salle à manger », et du maillage (voir figure 3.9) calculé par l'algorithme de radiosité hiérarchique : on voit que les frontières des ombres ont été correctement échantillonnées, mais que sur le mur, on observe un maillage très fin. On peut penser qu'il y a eu suréchantillonnage par rapport à la variation de la radiosité à cet endroit.

Ce raffinement excessif est produit par la nature même de l'oracle de raffinement utilisé ; le raffinement est uniquement basé sur la taille du facteur de forme, multiplié par la radiosité transportée. Or il peut arriver qu'on ait une valeur du facteur de forme précise bien qu'élevée. Dans l'idéal, il faudrait pouvoir mesurer l'erreur produite sur le facteur de forme lui-même, et baser l'oracle de raffinement sur cette erreur (voir chapitre 6).

Une mesure applicable simplement et sans calculer de nouvelles quantités (facteurs de forme ou leurs dérivées) est de mesurer *a posteriori* le bien fondé du raffinement décidé par l'oracle – quel qu'il soit – en calculant l'information apportée par ce raffinement.

Si l'on considère deux éléments p et q, qui interagissent. Le lien bi-directionnel qui les relie porte deux approximations de facteurs de forme, $F_{p,q}$ and $F_{q,p}$. Si nous raffinons ce lien en subdivisant p en éléments plus petits p_i, d'aire A_i (en utilisant par exemple un quadtree, ou un BSP-Tree), la définition même du facteur de forme :

$$F_{u,v} = \frac{1}{A_u} \int_{A_u} \int_{A_v} G(dA_u, dA_v)\, dA_u dA_v \tag{3.4}$$

où G est une fonction exprimant la géométrie relative des éléments de surface dA_u et

dA_v entraîne que les nouveaux facteurs de forme vérifient les relations :

$$F_{p,q} = \frac{1}{A_p}\left(\sum_i A_i F_{p_i,q}\right) \tag{3.5}$$

$$F_{q,p} = \sum_i F_{q,p_i} \tag{3.6}$$

Naturellement, ces relations ne concernent que les valeurs exactes des facteurs de forme. Cependant, elles peuvent être utilisées pour comparer, après raffinement, les nouvelles valeurs du facteur de forme trouvées avec les anciennes valeurs connues, et ainsi déterminer si le raffinement était justifié ou non, au vu de l'information qu'il apporte.

Ainsi, si les nouveaux $F_{p_i,q}$ sont peu différents les uns des autres, si la somme des F_{q,p_i} est proche de l'ancien $F_{q,p}$, et si de plus la moyenne des $F_{p_i,q}$ est proche de l'ancien $F_{p,q}$, alors ce raffinement n'avait pas d'utilité. Dans ce cas, on annule l'interaction, et on force p et q à interagir au niveau actuel, puisqu'il semble que les approximations du facteur de forme soit suffisamment précises.

Évidemment, cet algorithme de réduction du nombre de liens ne peut s'appliquer de façon sûre que dans des situations de visibilité totale entre les facettes : sinon, une apparente égalité des facteurs de forme pourrait n'être due qu'au hasard des différents obstacles.

Si l'on choisit de noter $F_{u,v}$ notre estimation courante du facteur de forme entre deux éléments, u et v, et si nous voulons raffiner un élément p en éléments p_i, nous notons :

$$F^{\min}_{p,q} = \min_i(F_{p_i,q}) \qquad F^{\min}_{q,p} = \min_i(F_{q,p_i})$$

$$F^{\max}_{p,q} = \max_i(F_{p_i,q}) \qquad F^{\max}_{q,p} = \max_i(F_{q,p_i})$$

$$F'_{p,q} = \tfrac{1}{A_p}\left(\textstyle\sum_i A_i F_{p_i,q}\right) \qquad F'_{q,p} = \textstyle\sum_i F_{q,p_i}$$

et l'on raffine effectivement p si l'une des conditions suivantes est vraie :

$$\frac{F^{\max}_{p,q} - F^{\min}_{p,q}}{F^{\max}_{p,q}} \geq \varepsilon_{\text{reduce}} \qquad \frac{F^{\max}_{q,p} - F^{\min}_{q,p}}{F^{\max}_{q,p}} \geq \varepsilon_{\text{reduce}}$$

$$\frac{\left|F'_{p,q} - F_{p,q}\right|}{F'_{p,q}} \geq \varepsilon_{\text{reduce}} \qquad \frac{\left|F'_{q,p} - F_{q,p}\right|}{F'_{q,p}} \geq \varepsilon_{\text{reduce}}$$

La décision d'annuler le raffinement d'un lien est basée uniquement sur des facteurs géométriques. Dès lors, elle peut être considérée comme permanente : si le raffinement a été refusé une fois, il le sera à toutes les étapes successives. On peut donc marquer le lien comme «à ne pas raffiner» pour toute la suite des calculs.

Naturellement, il est aussi possible d'utiliser un critère de raffinement du style *BF* : on multiplie les différences calculées par la radiosité de l'émetteur, et on compare le résultat avec le seuil $\varepsilon_{\text{reduce}}$. Dans ce cas, il faut tester à nouveau à chaque itération s'il faut ou non raffiner le lien concerné, puisque sa radiosité a pu être modifiée.

Pour vérifier si le raffinement d'un lien était justifié, il a fallu calculer les facteurs de forme approchés pour tous les enfants du patch p. Donc, si cette méthode ne permet d'éviter qu'un seul niveau de raffinement, le temps de calcul sera sensiblement égal à ce qu'il serait sans utiliser notre critère de simplification. En revanche, le gain en mémoire

sera toujours présent. Mais si la réduction du nombre de liens se produit suffisamment en amont, plusieurs étapes de raffinement peuvent être ainsi évitées.

Une implémentation de cet algorithme a montré une réduction significative du nombre d'éléments dans les quadtrees, ainsi que du nombre de liens (voir figure 3.6), et une réduction légèrement plus faible du temps de calcul, à cause du coût des facteurs de forme supplémentaires (voir figure 3.8).

3.7 Résultats

3.7.1 Évaluation paresseuse

La figure 3.10 montre la même scène que la figure 3.9, mais les calculs ont été effectués en utilisant la stratégie décrite dans la section 3.5. On remarque que les deux figures sont impossibles à distinguer à l'œil nu. La figure 3.10 montre également les différences entre les deux images. L'échelle des couleurs pour la figure des différences a été multipliée par huit par rapport aux images calculées. Les zones noires correspondent à pas d'erreur du tout, les zones blanches les plus claires de l'image à une erreur de l'ordre de 10 %.

3.7.2 Réduction du nombre de liens

Pour mesurer l'efficacité du critère de réduction des liens, nous avons calculé le rapport du nombre d'éléments dans la hiérarchie avec ce critère et du nombre d'élément avec l'algorithme original. La courbe de la figure 3.6 montre l'évolution de ce rapport avec le nombre d'itérations. Le nombre moyen de facettes a été en gros divisé par deux pour toutes les scènes de tests. La figure 3.7 montre le même rapport pour le nombre de liens. Cette réduction générale du nombre d'objets nécessaires pour résoudre le problème d'éclairage entraîne une réduction similaire de la place mémoire nécessaire pour l'algorithme, rendant ainsi plus accessibles les calculs sur des scènes avec de nombreux polygones.

La figure 3.8 illustre le rapport entre les temps de calcul de l'ancien algorithme et ceux avec le nouveau critère de raffinement. La réduction du nombre de liens *a posteriori* a un effet très visible sur le temps de calcul, avec des accélérations supérieures à 50%.

La figure 3.11 montre les calculs effectués par l'algorithme avec réduction du nombre de liens. On voit une nette diminution du nombre de facettes, alors que l'image finale est très semblable à l'image originale (figure 3.9). La figure 3.12 montre la différence entre les deux images, avec la même échelle de couleurs que la figure 3.10.

Comme une partie des différences est due à l'algorithme d'évaluation paresseuse des liens de haut niveau, la figure 3.12 montre aussi la différence entre l'algorithme avec évaluation paresseuse et l'algorithme avec évaluation paresseuse *et* réduction des liens. On peut voir que l'erreur introduite par la réduction des liens est très faible.

3.7.3 Temps total de calcul

Les temps de calcul sont présentés dans le tableau 3.3. Comme prévu, l'algorithme a été accéléré de façon significative, en particulier pour des scènes complexes. Pour toutes nos scènes de test, dix itérations avec une évaluation paresseuse des liens de haut niveau ont pris moins de temps que la première itération avec l'algorithme original. En

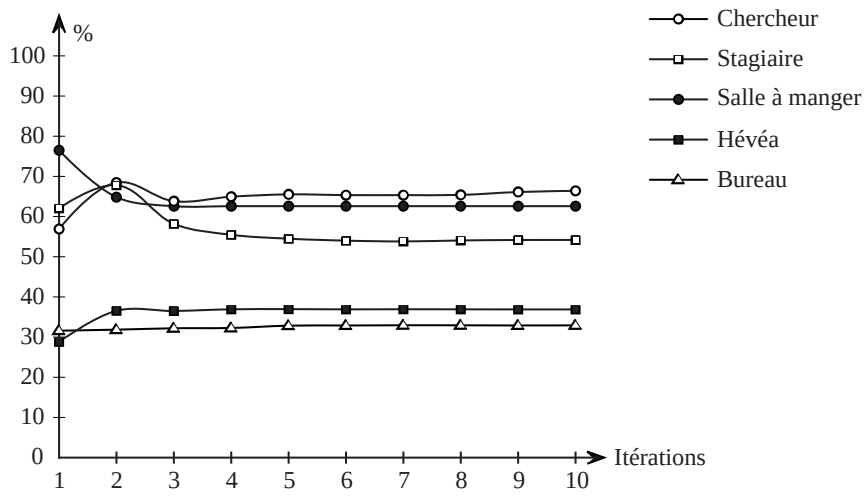


FIG. 3.6 – *Pourcentage du nombre d'éléments conservés après la réduction.*

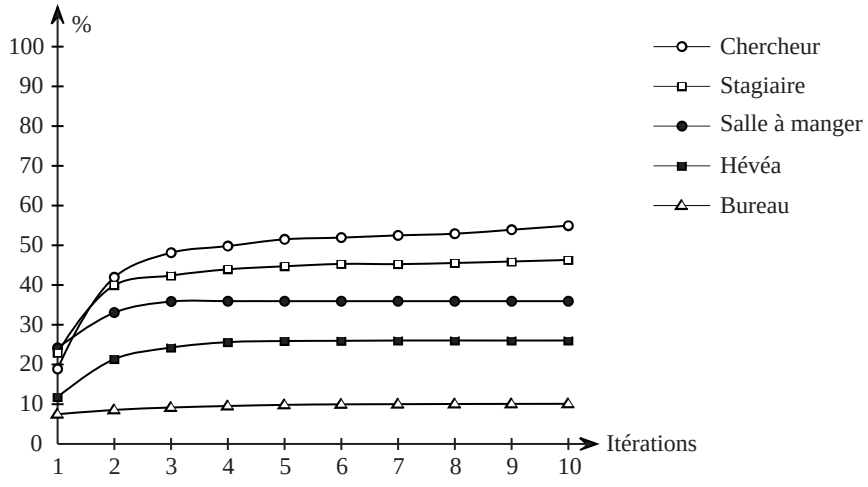


FIG. 3.7 – *Pourcentage du nombre de liens conservés après la réduction.*

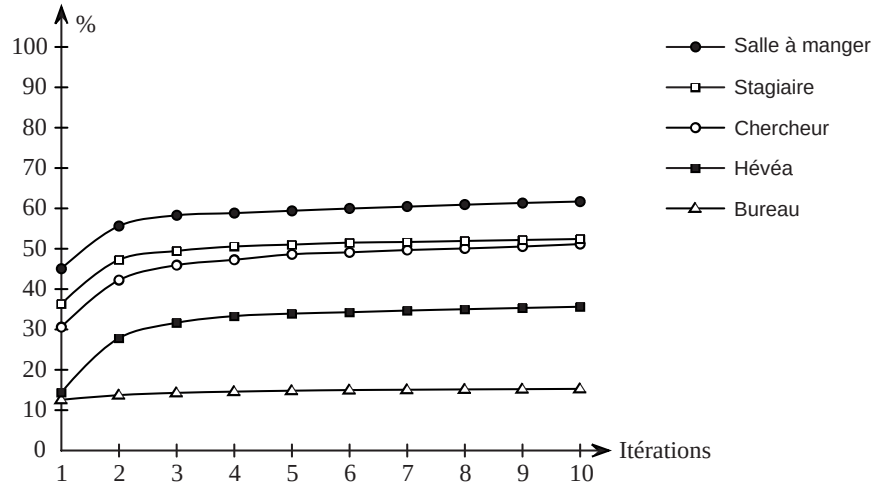


FIG. 3.8 – *Rapport des temps de calcul avec et sans la réduction.*

utilisant conjointement la réduction du nombre de liens et l'évaluation paresseuse, on arrive à produire une image utilisable en quelques minutes, même pour les scènes les plus complexes de notre test.

TAB. 3.3 – *Temps nécessaire pour dix itérations (et temps nécessaire pour produire la première image).*

Nom	*n*	Algorithme Original		Évaluation paresseuse …		et Réduction	
Bureau	170	301 s	(242 s)	287 s	(234 s)	43 s	(30 s)
Salle à Manger	402	4824 s	(4191 s)	4051 s	(3911 s)	657 s	(552 s)
Chercheur	1006	587 s	(378 s)	377 s	(191 s)	193 s	(59 s)
Stagiaire	1647	1017 s	(752 s)	514 s	(277 s)	270 s	(101 s)
Hévéa	2355	4253 s	(2331 s)	1526 s	(847 s)	543 s	(122 s)

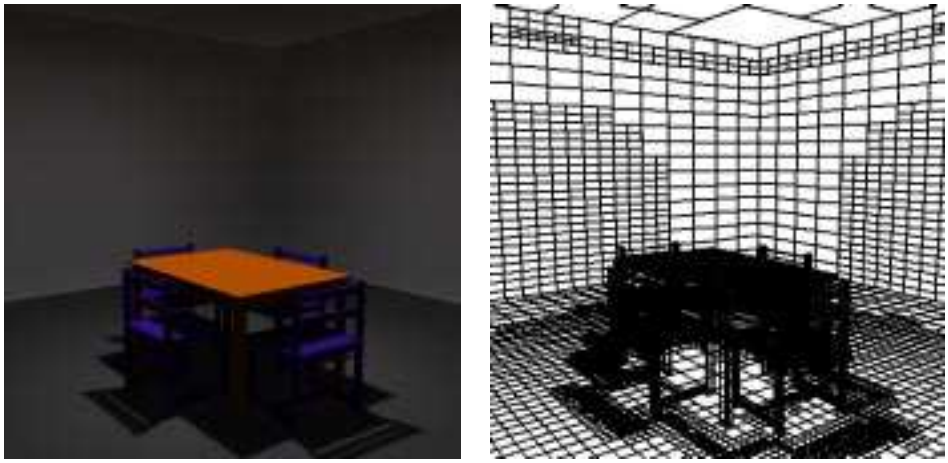


FIG. 3.9 – *La salle à manger avec l'algorithme original et le maillage produit.*

FIG. 3.10 – *La salle à manger avec l'évaluation paresseuse, et la différence avec 3.9.*

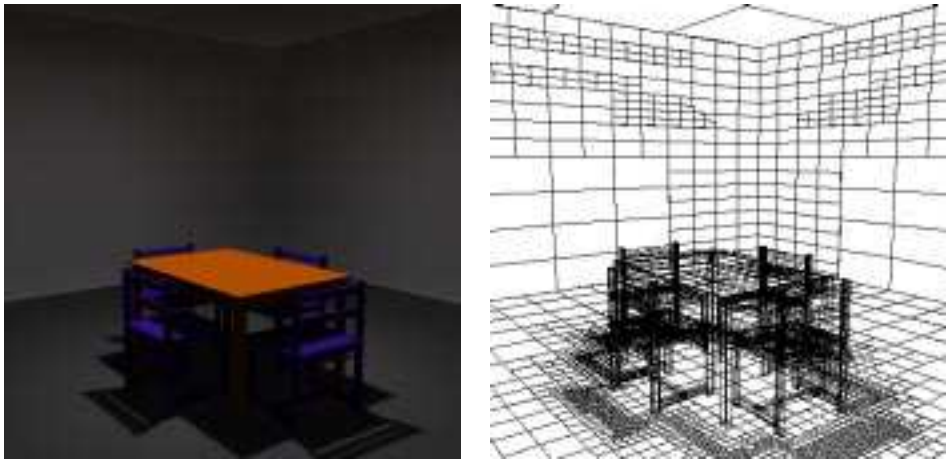


FIG. 3.11 – *La salle à manger avec réduction du nombre de liens, et le maillage produit.*

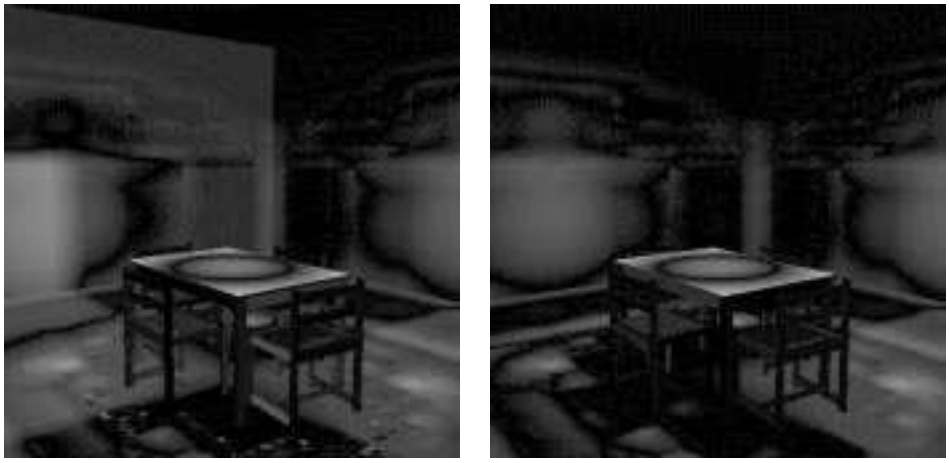


FIG. 3.12 – *Différence entre 3.11 et 3.9, à gauche, entre 3.11 et 3.10 à droite.*

FIG. 3.13 – *Bureau et Chercheur : deux modèles de bureaux.*

FIG. 3.14 – *Stagiaire et Hévéa : deux scènes plus complexes.*

Deuxième partie

Le contrôle de l’erreur en radiosité hiérarchique

4.

Notions sur les fonctions de plusieurs variables utiles pour les calculs de radiosité

SUR chacun des objets composant la scène, la radiosité est une fonction qui dépend de la position du point considéré. L'étude des variations est importante pour juger de la qualité de la discrétisation effectuée. Comme la radiosité est définie sur un objet tri-dimensionnel, on ne peut pas utiliser la notion classique de dérivée que l'on utilise pour les fonctions normales, de $\mathbb{R}$ dans $\mathbb{R}$. Cependant, les notions de dérivée, et même de dérivées successives, s'étendent parfaitement, comme nous allons le voir, aux fonctions dépendant de plusieurs paramètres.

De telles notions apportent toujours une information précieuse sur le comportement local de la fonction étudiée : ses variations, sa convexité et donc sa position par rapport au plan tangent, etc. De plus, dans certains cas, les valeurs de ces dérivées en quelques points permettent de déduire des informations valables sur tout un intervalle, ou sur tout un polygone.

4.1 Travaux antérieurs

L'idée d'utiliser les dérivées de la radiosité pour améliorer la qualité de la modélisation n'est pas nouvelle ; on peut voir, par exemple, les travaux de Vedel [44] et ceux de Ward [45], publiés en 1992, sur différentes manières d'utiliser le gradient de la radiosité et de l'approximer. Voir également le livre de Cohen [7], qui contient un très bon résumé des différents critères de raffinement possibles, basés sur les dérivées de la radiosité.

Le premier calcul exact des dérivées de la radiosité a été effectué par Arvo, en 1994 [3], pour des émetteurs constants, en présence d'obstacles. La quantité calculée n'est pas directement le gradient de la radiosité sur le récepteur, mais le jacobien du vecteur d'irradiance, ce qui est plus général.

Une idée plus subtile est celle de Michelin, en 1994 [2, 29], qui utilise, non pas les dérivées de la radiosité, mais la valeur exacte des dérivées du facteur de forme, pour en déduire rapidement une approximation du facteur de forme dans toutes les directions.

Pour ses premières sections, ce chapitre rappelle quelques résultats bien connus sur les extensions de la notion de dérivation aux fonctions de plusieurs variables (gradient,

jacobien, hessien, divergence, rotationnel). J'emprunte les notations et formules utilisées à Schwartz [35] ainsi qu'à des sources personnelles [46, 31]. Les sections suivantes montrent comment il est possible de calculer les dérivées successives de la radiosité (section 4.6 et 4.7), quelle que soit la configuration. Enfin, nous verrons, à la section 4.8, comment l'utilisation formelle des opérateurs de dérivation (gradient, rotationnel) permet de simplifier l'écriture de la radiosité, là encore dans tous les cas de figure.

4.2 La notion de dérivée en dimension n

La dérivée est définie pour les fonctions f de $\mathbb{R}$ dans $\mathbb{R}$ comme la limite, lorsqu'elle existe et qu'elle est unique de la fonction :

$$\frac{f(x+h)-f(x)}{h}$$

La notion s'étend naturellement aux fonction de plusieurs variables : on dit que la fonction $f(\vec{u})$ admet une dérivée dans la direction $\vec{v}$ lorsque

$$\frac{f(\vec{u}+h\vec{v})-f(\vec{u})}{h}$$

admet une limite et qu'elle est unique.

Un cas naturellement intéressant se présente lorsque la fonction $f(\vec{u})$ admet une limite dans toutes les directions $\vec{v}$. Un cas encore plus intéressant est lorsque toutes ces dérivées sont liées entre elles. Notamment lorsqu'elles sont toutes déductibles de $\vec{v}$ par une fonction linéaire. On dit alors que la fonction $f(\vec{u})$ est *différentiable*.

Si $f(\vec{u})$ est de $\mathbb{R}^n$ dans $\mathbb{R}^p$, la dérivée de $f(\vec{u})$ dans la direction $\vec{v}$ est un vecteur de $\mathbb{R}^p$. Si $f(\vec{u})$ est différentiable, cette dérivée se déduit de $\vec{v}$ par une transformation linéaire. $\vec{v}$ est un vecteur de $\mathbb{R}^n$, donc cette transformation linéaire ne peut être qu'une transformation linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. Un telle transformation linéaire a une notation naturelle sous forme d'une matrice à n lignes et p colonnes. On appelle cette matrice le *Jacobien* de f en $\vec{u}$.

Le Jacobien, lorsqu'il existe, correspond à la matrice des dérivées partielles de f. Ainsi, l'élément $i \times j$ du Jacobien est la dérivée suivant la i$^{\text{e}}$ coordonnée de la j$^{\text{e}}$ coordonnée de la fonction f. Le Jacobien d'une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}^p$ s'écrirait donc :

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial f_1(\vec{r})}{\partial x_1} & \cdots & \frac{\partial f_1(\vec{r})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(\vec{r})}{\partial x_1} & \cdots & \frac{\partial f_p(\vec{r})}{\partial x_n} \end{bmatrix}$$

Le Jacobien est une généralisation de la notion de dérivée qui conserve les propriétés habituelles de la dérivée : ainsi, si la variation d'une fonction dans une direction $\vec{v}$ peut être approximée en utilisant le Jacobien $\boldsymbol{J}$:

$$f(\vec{u}+h\vec{v}) = f(\vec{u}) + h\boldsymbol{J}\vec{v} + O(h^2)$$

L'application :

$$g(\vec{v}) = f(\vec{u}) + \boldsymbol{J}\vec{v}$$

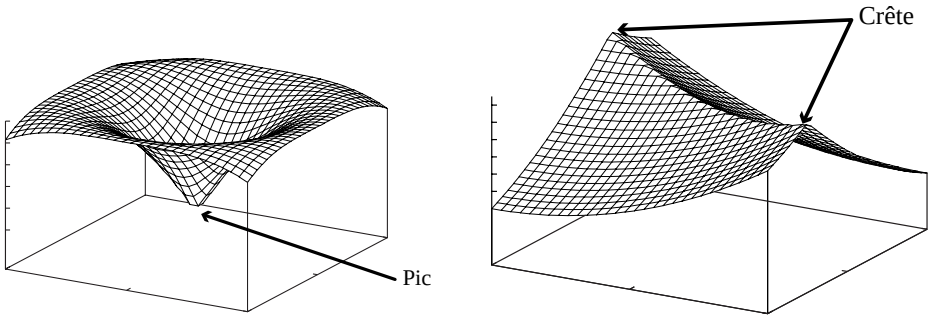


FIG. 4.1 – *Exemples de fonctions non-différentiables.*

est appelée application linéaire tangente à f. C'est une application linéaire, et sa différence avec f est en $O(\|v\|^2)$, comme on peut le voir plus loin, par exemple figure 4.2, ou aussi figure 5.6.

Ainsi, si une fonction est différentiable en un point, sa représentation est «lisse». Inversement, si une fonction n'est pas différentiable en un point, alors elle a généralement en ce point un «pic» ou une «crête» (voir figure 4.1).

4.3 Quelques cas particuliers intéressants : gradient, rotationnel et divergence

Lorsque la fonction considérée est à valeur dans $\mathbb{R}$: $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, le Jacobien se réduit à une matrice à n lignes et 1 colonne, donc à un vecteur de $\mathbb{R}^n$. Notons ce vecteur $\vec{G}$. On a alors :

$$f(\vec{u}+h\vec{v}) = f(\vec{u}) + h\vec{G}\cdot\vec{v} + O(h^2)$$

C'est-à-dire que le produit matriciel s'est transformé en produit scalaire de deux vecteurs. Le vecteur $\vec{G}$ est appelé *gradient* de f au point $\vec{u}$.

Le gradient est donc l'expression de la différentielle pour les fonctions d'un espace vectoriel à valeur dans $\mathbb{R}$. On a l'habitude de le noter $\nabla f(\vec{u})$. L'application linéaire tangente devient dans ce cas :

$$g(\vec{v}) = f(\vec{u}) + \vec{v}\cdot\vec{G}$$

g définit le sous-espace tangent à la fonction, donc le plan tangent lorsque $n = 2$ (voir figure 4.2). Le gradient définit à la fois le plan tangent à la fonction et la ligne de plus grande pente : c'est dans la direction du gradient que la fonction varie le plus.

Lorsque l'on a une fonction $\vec{f}$ de $\mathbb{R}^3$ dans $\mathbb{R}^3$, on introduit deux quantités supplémentaires : le rotationnel, noté $\nabla\times\vec{f}$ et la divergence, notée $\nabla\cdot\vec{f}$:

$$\nabla\times\vec{f}(\vec{u}) = \begin{pmatrix} \frac{\partial f_y(\vec{u})}{\partial z} - \frac{\partial f_z(\vec{u})}{\partial y} \\ \frac{\partial f_z(\vec{u})}{\partial x} - \frac{\partial f_x(\vec{u})}{\partial z} \\ \frac{\partial f_x(\vec{u})}{\partial y} - \frac{\partial f_y(\vec{u})}{\partial x} \end{pmatrix}$$

$$\nabla\cdot\vec{f}(\vec{u}) = \frac{\partial f_x(\vec{u})}{\partial x} + \frac{\partial f_y(\vec{u})}{\partial y} + \frac{\partial f_z(\vec{u})}{\partial z}$$

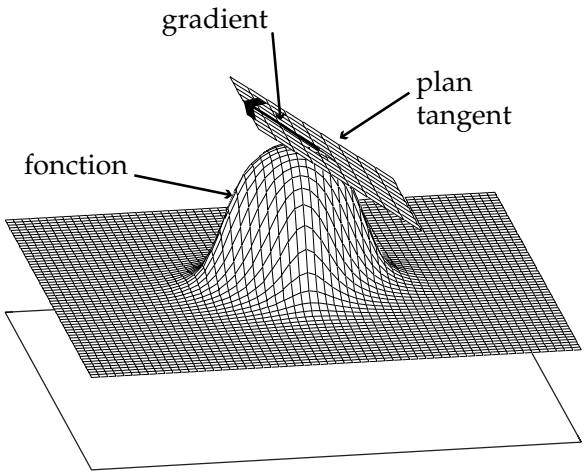


FIG. 4.2 – *Le gradient définit un plan tangent.*

Ces deux quantités ont leur origine en mécanique des milieux continus, où elles servent à exprimer les variations d'un milieu, d'un fluide : le rotationnel exprime combien les particules de fluide tournent autour d'un point, la divergence exprime la variation de volume des particules de fluide.

En ce qui nous concerne, les expressions du rotationnel et de la divergence n'ont plus de significations physiques simples dans les expressions que nous manipulons. Néanmoins, elles restent des outils excellents pour manipuler les formules comprenant des fonctions géométriques.

En mécanique des fluides, une fonction $f(\vec{r})$ de $\mathbb{R}^3$ dans $\mathbb{R}$, s'appelle un *champ*, ou un champ scalaire. Une fonction $\vec{f}(\vec{r})$ de $\mathbb{R}^3$ dans $\mathbb{R}^3$ s'appelle un champ vectoriel.

Lorsqu'un champ vectoriel a une divergence nulle, alors il peut s'exprimer comme le rotationnel d'un autre champ vectoriel :

$$\nabla \cdot \vec{f} = 0 \Longrightarrow \exists \vec{V}, \vec{f} = \nabla \times \vec{V}$$

De la même manière, lorsqu'un champ vectoriel a un rotationnel nul, alors il peut s'exprimer comme le gradient d'un champ scalaire.

$$\nabla \times \vec{f} = 0 \Longrightarrow \exists U, \vec{f} = \nabla U$$

Lorsqu'on a affaire à des intégrales de champs vectoriels ou scalaires, il peut-être utile de connaître les théorèmes de Stokes et d'Ostrogradsky. Le théorème de Stokes dit que le flux du rotationnel d'un champ vectoriel à travers une surface est égal à la circulation de ce champ sur le contour de la surface :

$$\oint_C \vec{a} \cdot d\vec{x} = \int_S (\nabla \times \vec{a}) \cdot d\vec{S} \tag{4.1}$$

Le théorème d'Ostrogradsky est assez proche :

$$\oint_C U d\vec{x} = -\int_S \nabla U \times d\vec{S} \tag{4.2}$$

En fait, ces deux théorèmes sont équivalents l'un à l'autre. Il est très simple de démontrer la formule de Stokes à partir de celle d'Ostrogradsky, et réciproquement. Nous verrons plus loin (section 4.8 et chapitre 10) comment ces deux formules peuvent permettre, dans certains cas, de simplifier l'expression de la radiosité.

4.4 Les dérivées successives

Pour une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}^p$ le Jacobien, $\boldsymbol{J}$ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

Si f est dérivable en tout point de son ensemble de définition, le Jacobien peut être défini comme une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^p \times \mathbb{R}^n$.

De la même manière que f, $\boldsymbol{J}$ peut admettre une dérivée suivant un vecteur, voire être différentiable. La différentielle d'une différentielle est une application linéaire $\boldsymbol{H}$ qui à tout vecteur $\vec{u}$ associe une application linéaire $\boldsymbol{H}(\vec{u})$, approximation de la variation de $\boldsymbol{J}$ dans la direction $\vec{u}$. De telles applications sont dites bilinéaires.

On a ainsi :

$$\boldsymbol{J}(\vec{v}+h\vec{u}) = \boldsymbol{J}(\vec{v}) + h\boldsymbol{H}\vec{u} + O(h^2)$$

Donc, si une fonction est deux fois différentiable en un point donné, on peut approximer sa variation dans une direction $\vec{v}$ par une formule du second ordre :

$$f(\vec{u}+h\vec{v}) = f(\vec{u}) + h\boldsymbol{J}\vec{v} + h^2(\boldsymbol{H}\vec{v})\vec{v} + O(h^3)$$

Pour éviter la redondance de la notation $(\boldsymbol{H}\vec{v})\vec{v}$, on préfère souvent noter les applications bilinéaires sous la forme : ${}^{\mathrm{t}}\vec{v}\boldsymbol{H}\vec{v}$ (naturellement, l'expression matricielle de $\boldsymbol{H}$ n'est alors plus la même). Dans ce cas,

$$f(\vec{u}+h\vec{v}) = f(\vec{u}) + h\boldsymbol{J}\vec{v} + h^2{}^{\mathrm{t}}\vec{v}\boldsymbol{H}\vec{v} + O(h^3)$$

L'application bilinéaire $\boldsymbol{H}$ se nomme le *Hessien* de f au point u, tout comme $\boldsymbol{J}$ s'appelle le Jacobien.

Dans le cas d'une fonction d'un espace vectoriel à valeurs réelles, $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, le Jacobien se réduit à un vecteur ; le Hessien est une application bilinéaire de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$. Ces applications bilinéaires ont une notation matricielle : l'élément $i \times j$ de la matrice du Hessien est la dérivée $\frac{\partial^2 f}{\partial x_i \partial x_j}$:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f(\vec{r})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\vec{r})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\vec{r})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\vec{r})}{\partial x_n \partial x_n} \end{bmatrix}$$

Lorsque le Hessien existe, il est nécessairement symétrique, à cause de la formule de Schwartz :

$$\frac{\partial^2 f(\vec{r})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\vec{r})}{\partial x_j \partial x_i}$$

La principale utilisation des dérivées successives est pour encadrer les fonctions étudiées, en identifiant leurs maximums et leur minimums. Il est évident qu'en un maximum ou un minimum, si la fonction est différentiable, alors sa différentielle est nulle.

Cependant, cette condition n'est pas suffisante : il existe des points où la différentielle est nulle sans que la fonction ait d'extremum.

En revanche, si la différentielle est nulle et si le Hessien est défini positif, alors f admet en ce point un minimum local. Si le Hessien est défini négatif, alors f admet un maximum local.

Si le Hessien n'est ni défini-positif, ni défini-négatif, alors on ne peut pas conclure.

Pour étudier la position de la fonction par rapport à son plan tangent au point $\vec{r}$, on étudie en fait :

$$g(\vec{u}) = f(\vec{r}+\vec{u}) - \boldsymbol{J}\vec{u}$$

g est définie, deux fois différentiable. Sa différentielle en 0 est nulle. Son Hessien en 0 est celui de f. Si g admet un minimum relatif en 0, alors f est au dessus de son sous-espace tangent en 0. Si g admet un maximum relatif en 0, alors f est en dessous de son sous-espace tangent en 0.

Donc :

– Si le Hessien de f est défini-positif au point $\vec{r}$, alors f est au dessus de son sous-espace tangent sur un voisinage du point $\vec{r}$:

$$f(\vec{r}+\vec{u}) \geq f(\vec{r}) + \boldsymbol{J}\vec{u}$$

– Si le Hessien de f est défini-négatif au point $\vec{r}$, alors f est en dessous de son sous-espace tangent sur un voisinage du point $\vec{r}$ (voir figure 4.3):

$$f(\vec{r}+\vec{u}) \leq f(\vec{r}) + \boldsymbol{J}\vec{u}$$

– Si le Hessien de f n'est pas défini, alors on ne peut pas conclure. En général, la fonction traverse son sous-espace tangent suivant deux droites (voir figure 4.4).

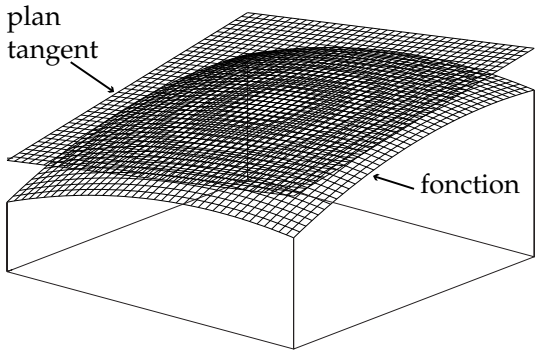


FIG. 4.3 – *Cas d'un point où le Hessien est défini-négatif.*

Des conditions définies seulement au voisinage d'un point n'ont pas vraiment d'utilité pour une étude globale de la fonction. Heureusement, cette propriété se généralise :

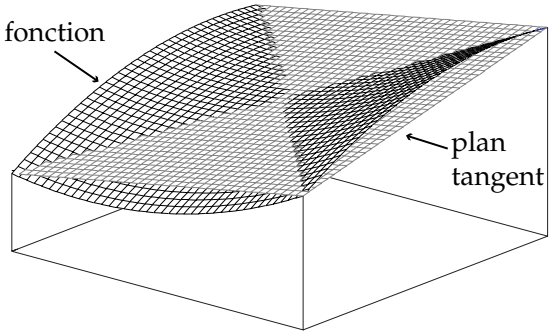


FIG. 4.4 – *Cas d'un point où le Hessien n'est pas défini.*

s'il existe un convexe sur lequel le Hessien de f soit défini-positif, alors sur ce convexe f est au dessus de tous ses plans tangents. Mieux encore, sur tout triangle inclus dans le convexe, f est en dessous du plan sécant défini par ses valeurs aux sommets du triangle. On parle d'une fonction convexe. Voir figure 4.5 un exemple de fonction convexe.

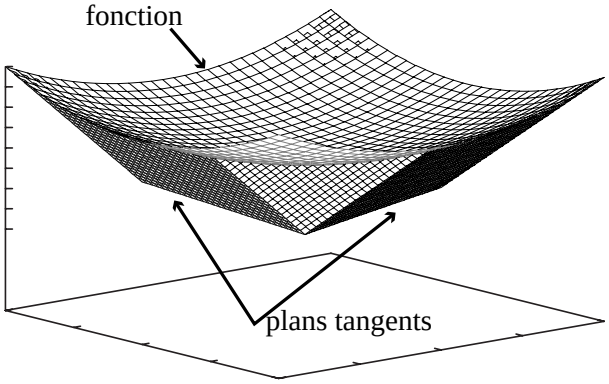


FIG. 4.5 – *Une fonction convexe est au dessus de ses plans tangents.*

Inversement, s'il existe un convexe sur lequel le Hessien de f soit défini-négatif, alors sur ce convexe f est en dessous de tous ses plans tangents, et au dessus de ses plans sécants. On parle d'une fonction concave.

4.5 Cas particulier des fonctions de deux variables

Le cas auquel nous nous intéressons généralement est celui de l'interaction d'objets surfaciques, et non volumiques. Dans ce cas, deux variables suffisent pour désigner tous les points de la surface. Sur un objet donné, nous pouvons donc considérer la radiosité comme une fonction de deux variables, u et v.

Dans ce cas, le Jacobien s'exprime comme un vecteur à deux coordonnées :

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial f(\vec{r})}{\partial u} \\ \frac{\partial f(\vec{r})}{\partial v} \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix}$$

Et le Hessien[1], comme une matrice 2×2 :

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f(\vec{r})}{\partial u^2} & \frac{\partial^2 f(\vec{r})}{\partial u \partial v} \\ \frac{\partial^2 f(\vec{r})}{\partial u \partial v} & \frac{\partial^2 f(\vec{r})}{\partial v^2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} r & s \\ s & t \end{bmatrix}$$

On distingue alors quatre cas :

1. $rt - s^2 > 0$ et $r < 0$: $\boldsymbol{H}$ est définie négative, f est concave, donc au dessous de son plan tangent ;
2. $rt - s^2 > 0$ et $r > 0$: $\boldsymbol{H}$ est définie positive, f est convexe, donc au dessus de son plan tangent ;
3. $rt - s^2 < 0$: on est en présence d'un col, et la fonction traverse son plan tangent suivant deux droites ;
4. $rt - s^2 = 0$: on ne peut conclure sans étudier les dérivées successives.

Naturellement, ces résultats sont indépendants du système de coordonnées orthonormées choisies pour paramétrer l'objet étudié. En effet, un changement de base de matrice $\boldsymbol{P}$ donne à f un nouveau Jacobien $\boldsymbol{JP}$, et un nouveau Hessien ${}^t\boldsymbol{PHP}$. Si $\boldsymbol{P}$ est orthonormée, ${}^t\boldsymbol{P} = \boldsymbol{P}^{-1}$. Or $rt - s^2$ est le déterminant de $\boldsymbol{H}$, qui est évidemment invariant par changement de base. Si $rt - s^2$ est positif, r et t sont de même signe. Ce signe est aussi celui des deux valeurs propres de $\boldsymbol{H}$.

4.6 Application à la radiosité

La radiosité est définie comme la solution de l'équation intégrale 2.11 :

$$B(x) = E(x) + \rho_d(x) \int_{y \in S} B(y) \frac{\cos\theta \cos\theta'}{\pi r^2} V(x,y) dy$$

Ou encore, en utilisant l'opérateur intégral $\mathcal{H}$,

$$B(x) = E(x) + \rho_d(x) \left(\mathcal{H}B\right)(x)$$

Cette équation ne définit $B(x)$ qu'en fonction de $B(y)$. On ne peut guère en tirer une expression des dérivées de $B(x)$ que sous forme de solutions d'une autre équation intégrale.

Cependant, on peut *singulariser* un instant donné du processus de résolution itérative. À cet instant, on a une connaissance approximative de la radiosité $B(x)$. Cette

1. Ces notations (p,q,r,s,t) sont des standards pour les fonctions de deux variables. Voyez notamment Schwartz, 1992 [35]

connaissance de la radiosité dépend naturellement d'une certaine discrétisation du problème – quelles que soient les fonctions de base que l'on utilise (linéaires, polynomiales, ondelettes). En ce qui concerne la décomposition de la radiosité suivant certaines fonctions de base, on peut consulter Troutman, 1993 [43], Zatz, 1993 [48] et Schröder, 1994 [32]. On peut alors vouloir accéder à une approximation des dérivées de la radiosité – sous les mêmes hypothèses de discrétisation, bien sûr.

La dérivation est un opérateur linéaire : la dérivée de $f+g$ est égale à la dérivée de f plus la dérivée de g. Le transport de lumière est également un processus linéaire : la fonction de radiosité sur une surface est la somme des contributions de tous les émetteurs de la scène.

Pour connaître la dérivée de la radiosité en un endroit, il suffit donc de calculer la somme des dérivées des contributions de tous les émetteurs.

Or la dérivée d'une contribution est aisément calculable. Supposons un émetteur A_2 et un récepteur A_1 (voir figure 4.6). La radiosité en un point x sur A_1 est uniquement due à A_2 :

$$B(x) = E(x) + \frac{\rho}{\pi} \int_{A_2} B(y) V(x,y) \frac{\cos\theta_1 \cos\theta_2}{\|r_{12}\|^2} dA_2 \tag{4.3}$$

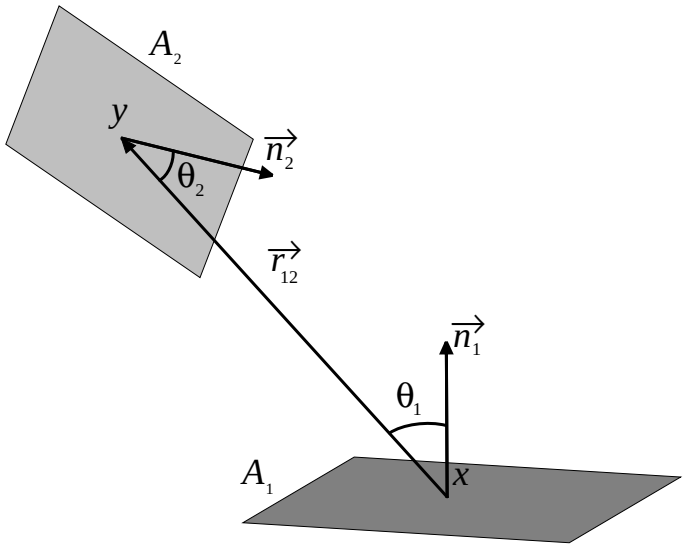


FIG. 4.6 – *Géométrie du problème.*

La fonction de visibilité $V(x,y)$ n'est pas dérivable, puisqu'elle ne peut prendre pour valeurs que 0 ou 1 : elle n'est même pas continue.

Mais on peut déplacer le problème : si l'intégrale de l'équation 4.3 porte sur la partie $A_2(x)$ de A_2 qui est visible de x, alors la définition de la radiosité sur A_1 peut s'écrire en fonction de :

$$B(x) = E(x) + \frac{\rho}{\pi} \int_{A_2(x)} B(y) \frac{\cos\theta_1 \cos\theta_2}{\|r_{12}\|^2} dA_2$$

En règle générale, $A_2(x)$ est une fonction dérivable de x, sauf sur les points du récepteur où deux arêtes de la scène sont vues comme la même arête, que ce soit des arêtes de l'émetteur ou des arêtes des obstacles.

Lorsque $A_2(x)$ est dérivable, la dérivée de cette expression de la radiosité suivant un vecteur $\vec{v}$ s'exprime comme :

$$\frac{dB}{d\vec{v}}(x) = \frac{dE}{d\vec{v}}(x) + \frac{\rho}{\pi}\left[\int_{A_2(x)} B(y)\frac{d}{d\vec{v}}\left(\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}\right)dA_2 + \frac{d}{d\vec{v}}\left(\int_{A_2(x)}\right)B(y)\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}dA_2\right]$$

Autrement dit, la dérivée suivant un vecteur de l'intégrale sur une surface d'une fonction est la somme de l'intégrale sur la surface de la dérivée de cette fonction et de l'intégrale sur la variation de la surface de la fonction.

Ce qui peut se voir autrement en considérant que l'intégrale est un opérateur linéaire. Si j'appelle I l'opérateur linéaire :

$$I(x) : g(y) \mapsto \frac{\rho}{\pi}\int_{A_2(x)} g(y)dy$$

et $f(x,y)$ la fonction :

$$f : (x,y) \mapsto B(y)\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}$$

Alors il est clair que : $B(x) = E(x) + I(x)\,f(x,y)$, et donc que :

$$\frac{dB}{d\vec{v}}(x) = \frac{dE}{d\vec{v}}(x) + I(x)\frac{df}{d\vec{v}}(x,y) + \frac{dI}{d\vec{v}}f(x,y)$$

Si l'on s'intéresse au gradient de la radiosité, et si $\boldsymbol{J}(\int_{A_2(x)})$ désigne le jacobien de l'opérateur d'intégration sur $A_2(x)$, alors :

$$\nabla B(x) = \nabla E(x) + \frac{\rho}{\pi}\left[\int_{A_2(x)} B(y)\nabla\left(\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}\right)dA_2 + \boldsymbol{J}\left(\int_{A_2(x)}\right)B(y)\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}dA_2\right] \quad (4.4)$$

De même, si l'on s'intéresse au Hessien de $B(x)$, noté $\boldsymbol{H}_B(x)$, on a :

$$\boldsymbol{H}_B(x) = \boldsymbol{H}_E(x) + \frac{\rho}{\pi}\left[\int_{A_2(x)} \boldsymbol{H}_f(x,y)dA_2 + \boldsymbol{H}_{\int_{A_2(x)}} f(x,y)dA_2 + 2\boldsymbol{J}\left(\int_{A_2(x)}\right)B(y)\nabla f(x,y)dA_2\right] \quad (4.5)$$

Nous verrons plus loin (chapitres 8 et 9) un exemple de calcul effectif du Hessien.

La convexité de $B(x)$, qui dépend du déterminant du Hessien de $B(x)$, n'est pas linéaire. La somme de deux fonctions convexes est certes convexe, la somme de deux fonctions concaves est également concave, mais le déterminant du Hessien de $B(x)$, $rt - s^2$ n'est pas la somme des déterminants des opérateurs qui composent le Hessien.

4.7 Calcul des dérivées successives de la radiosité pour un émetteur quelconque

Les équations 4.4 et 4.5 se simplifient quelque peu lorsque l'émetteur et le récepteur sont en situation de visibilité totale, c'est à dire lorsqu'il n'y a pas d'obstacles entre eux. Dans ce cas, l'intégrale porte sur toute la surface de l'émetteur A_2, et ce terme ne varie pas en fonction de la position de x.

On a donc :

$$B(x) = E(x) + \frac{\rho}{\pi}\int_{A_2} B(y)\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}dA_2$$

Dans ces conditions, le gradient de la radiosité s'exprime comme :

$$\nabla B(x) = \nabla E(x) + \frac{\rho}{\pi}\int_{A_2} B(y)\nabla\left(\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}\right)dA_2$$

Le terme sous le gradient à l'intérieur de l'intégrale ne dépend pas de $B(y)$; on peut le dériver formellement :

$$\nabla B(x) = \nabla E(x) + \frac{\rho}{\pi}\int_{A_2} B(y)\left(\frac{(\vec{r}_{12}\cdot\vec{n}_1)n_2 + (\vec{r}_{12}\cdot\vec{n}_2)n_1 - 4\vec{r}_{12}\cos\theta_1\cos\theta_2}{\|r_{12}\|^4}\right)dA_2$$

L'expression du gradient obtenue ainsi permet de calculer le gradient en toutes circonstances.

De la même manière, on peut calculer le Hessien de B :

$$\boldsymbol{H}_B(x) = \boldsymbol{H}_E(x) + \frac{\rho}{\pi}\int_{A_2} B(y)\boldsymbol{H}\left(\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}\right)dA_2$$

Le calcul pratique d'un Hessien se fait simplement en calculant la dérivée seconde $d_2(\vec{u},\vec{v})$ suivant deux vecteurs, puis en exhibant un opérateur $\boldsymbol{H}$ tel que : $d_2(\vec{u},\vec{v}) = {}^t\vec{u}\boldsymbol{H}\vec{v}$. Dans notre cas, on aurait une combinaison linéaire de termes de la forme $(\vec{a}\cdot\vec{u})(\vec{b}\cdot\vec{v}) + (\vec{a}\cdot\vec{v})(\vec{b}\cdot\vec{u})$ et de termes en $\vec{u}\cdot\vec{v}$, les coefficients ne dépendant pas de $\vec{u}$, ni de $\vec{v}$. Ces deux termes se traitent aisément :

$$\begin{aligned}
\vec{u}\cdot\vec{v} &= {}^t\vec{u}\vec{v} = {}^t\vec{u}\boldsymbol{I}\vec{v}\\
(\vec{a}\cdot\vec{u})(\vec{b}\cdot\vec{v}) &= ({}^t\vec{u}\vec{a})({}^t\vec{b}\vec{v}) = {}^t\vec{u}(\vec{a}{}^t\vec{b})\vec{v}\\
\vec{a}{}^t\vec{b} &= \begin{pmatrix} a_x\\ a_y\\ a_z\end{pmatrix}(b_x b_y b_z) = \begin{pmatrix} a_xb_x & a_xb_y & a_xb_z\\ a_yb_x & a_yb_y & a_yb_z\\ a_zb_x & a_zb_y & a_zb_z\end{pmatrix}\\
(\vec{a}\cdot\vec{u})(\vec{b}\cdot\vec{v}) + (\vec{a}\cdot\vec{v})(\vec{b}\cdot\vec{u}) &= {}^t\vec{u}(\vec{a}{}^t\vec{b} + \vec{b}{}^t\vec{a})\vec{v} = {}^t\vec{u}\boldsymbol{Q}(\vec{a},\vec{b})\vec{v}
\end{aligned}$$

On choisit de noter $\boldsymbol{Q}(\vec{a},\vec{b})$ la matrice $\vec{a}{}^t\vec{b} + \vec{b}{}^t\vec{a}$ pour simplifier les écritures. On remarque que $\boldsymbol{Q}(\vec{a},\vec{b})$ est naturellement symétrique.

Dans notre cas, la formule complète du Hessien lorsque la visibilité est complète

s'écrit :

$$\begin{aligned}
\boldsymbol{H}_B(x) &= \boldsymbol{H}_E(x) \\
&+ \frac{4\rho}{\pi}\int_{A_2} B(y)\frac{(\vec{r}_{12}\cdot\vec{n}_1)(\vec{r}_{12}\cdot\vec{n}_2)}{\|r_{12}\|^8}(3\boldsymbol{Q}(\vec{r}_{12},\vec{r}_{12})-\boldsymbol{I})dA_2 \\
&- \frac{4\rho}{\pi}\boldsymbol{Q}\left(\int_{A_2} B(y)\frac{(\vec{r}_{12}\cdot\vec{n}_1)\vec{r}_{12}}{\|r_{12}\|^6}dA_2,\vec{n}_2\right) \\
&- \frac{4\rho}{\pi}\boldsymbol{Q}\left(\int_{A_2} B(y)\frac{(\vec{r}_{12}\cdot\vec{n}_2)\vec{r}_{12}}{\|r_{12}\|^6}dA_2,\vec{n}_1\right) \\
&+ \frac{\rho}{\pi}\left(\int_{A_2} B(y)\frac{2}{\|r_{12}\|^4}dA_2\right)\boldsymbol{Q}(\vec{n}_1,\vec{n}_2)
\end{aligned}$$

On notera que les deux dérivations formelles effectuées ici sont indépendantes de toute considération sur B. Elles peuvent donc être utilisées dans tous les cas pour calculer le gradient et le Hessien de la radiosité en un point. Par exemple, si l'on utilise une base quelconque de fonctions (linéaires, polynomiales, ondelettes), il suffira de calculer le gradient et le Hessien pour chacune des fonctions de base, éventuellement en faisant un calcul approché des intégrales.

On remarque aussi que dans la plupart des calculs, il est possible de calculer d'abord une simple intégrale vectorielle, avant d'en tirer la forme matricielle du Hessien. La principale difficulté est le calcul de l'intégrale avec $\boldsymbol{Q}(\vec{r}_{12},\vec{r}_{12}) = 2\vec{r}_{12}{}^{\mathrm{t}}\vec{r}_{12}$. Si on a une paramétrisation de $\vec{r}_{12}$ sur A_2 :

$$\vec{r}_{12} = \vec{r}_0 + u\vec{a} + v\vec{b}$$

alors,

$$\vec{r}_{12}{}^{\mathrm{t}}\vec{r}_{12} = \vec{r}_0{}^{\mathrm{t}}\vec{r}_0 + u\boldsymbol{Q}(\vec{r}_0,\vec{a}) + v\boldsymbol{Q}(\vec{r}_0,b) + uv\boldsymbol{Q}(\vec{a},\vec{b}) + u^2\vec{a}^{\mathrm{t}}\vec{a} + v^2\vec{b}^{\mathrm{t}}\vec{b}$$

et on est ramené à des calculs d'intégrales vectorielles plus classiques.

Bien que le Hessien soit un opérateur matriciel 2×2, nous l'avons ici comme somme de matrices 3×3, à cause de la nature intrinsèquement tridimensionnelle des objets manipulés. Il est naturellement possible d'en déduire la valeur de la dérivée seconde suivant une direction donnée, en appliquant le Hessien à cette direction. De même, on peut en tirer la concavité en utilisant une projection du Hessien sur le plan récepteur considéré, et en prenant le déterminant de cette projection.

4.8 Simplification de l'écriture de la radiosité

Les outils mathématiques que nous venons de voir peuvent également être utilisés pour simplifier l'écriture de la radiosité. Nous avons vu, à la section 2.4, que la radiosité en un point x d'un récepteur A_1, due à un émetteur A_2 (voir figure 4.6), pouvait s'exprimer comme une intégrale sur la surface de A_2 (équation 2.11) :

$$B(x) = E(x) + \frac{\rho}{\pi}\int_{A_2} B(y)\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}dA_2 \tag{4.6}$$

Cette expression peut se simplifier, comme on le sait depuis le XVIII$^{\mathrm{e}}$ siècle et les travaux de Lambert, lorsque la radiosité sur l'émetteur est constante (voir aussi Baum,

1989 [5] et Schröder, 1993 [33, 34]). Dans ce cas, l'équation 4.6 se réduit à une intégration sur les contours de l'émetteur :

$$B(x) = E(x) - B_0 \frac{\rho}{2\pi} \vec{n}_1 \cdot \oint_{\partial A_2} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 \tag{4.7}$$

Le passage de l'équation 4.6 à l'équation 4.7 se fait en utilisant le théorème de Stokes (voir équation 4.1).

En effet, l'équation 4.6 n'est autre que le flux d'un champ vectoriel :

$$\begin{aligned}
\int_{A_2} B(y) \frac{\cos\theta_1 \cos\theta_2}{\|r_{12}\|^2} dA_2 &= -\int_{A_2} B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)(\vec{r}_{12} \cdot \vec{n}_2)}{\|\vec{r}_{12}\|^4} dA_2 \\
&= -\int_{A_2} B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \cdot d\vec{A}_2 \\
&= \int_{A_2} \vec{F}(\vec{r}_{12}) \cdot d\vec{A}_2
\end{aligned}$$

Avec :

$$\vec{F}(\vec{r}_{12}) = -B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)\vec{r}_{12}}{\|r_{12}\|^4}$$

Lorsque la radiosité sur l'émetteur est supposée constante, c'est-à-dire lorsqu'on a discrétisé l'émetteur en facettes sur lesquelles la radiosité est supposée constante, ce champ vectoriel $\vec{F}$ peut s'exprimer comme le rotationnel d'un autre champ vectoriel. Il suffit alors d'appliquer le théorème de Stokes pour retrouver l'équation 4.7.

Cette simplification permet une expression du gradient et du Hessien de la radiosité également comme des intégrales sur les contours, ce qui rend possible leur calcul dans des temps raisonnables.

Habituellement, lorsque la radiosité sur l'émetteur n'est pas constante, on est réduit à approximer l'intégrale surfacique. Cependant, les intégrales curvilignes sont, en règle générale, à la fois plus faciles à calculer expressément et plus faciles à approximer. Pour cette raison, il serait bon d'avoir une expression de la radiosité sur le récepteur comme une intégrale de contour, même lorsque la radiosité n'est pas constante sur l'émetteur.

Pour pouvoir passer d'une intégrale surfacique à une intégrale de contours, il faut appliquer le théorème de Stokes. Et pour pouvoir appliquer le théorème de Stokes, il faut exprimer le champ vectoriel dont on prend le flux, $\vec{F}(\vec{r}_{12})$, comme le rotationnel d'un champ vectoriel.

Pour qu'il soit possible d'exprimer un champ vectoriel $\vec{V}(\vec{r})$, comme le rotationnel d'un autre champ vectoriel $\vec{U}(\vec{r})$, il faut et il suffit que la divergence du premier champ vectoriel soit nulle :

$$\nabla \cdot (\vec{V}(\vec{r})) = 0 \Longleftrightarrow \exists \vec{U}(\vec{r}), \vec{V}(\vec{r}) = \nabla \times (\vec{U}(\vec{r}))$$

Calculons donc la divergence du champ vectoriel $\vec{F}$ dont on calcule le flux à travers

A_2 :

$$\begin{aligned}
\nabla \cdot \vec{F}(\vec{r}_{12}) &= \nabla \cdot \left(-B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)}{\|r_{12}\|^4} \vec{r}_{12} \right) \\
&= -\nabla \left(B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)}{\|r_{12}\|^4} \right) \cdot \vec{r}_{12} - 3 \left(B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)}{\|r_{12}\|^4} \right) \\
\nabla \left(B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)}{\|r_{12}\|^4} \right) &= \nabla B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)}{\|r_{12}\|^4} - 3B(y) \frac{\vec{n}_1}{\|r_{12}\|^4} \\
\nabla \cdot \vec{F}(\vec{r}_{12}) &= -\nabla B(y) \cdot \vec{r}_{12} \frac{(\vec{r}_{12} \cdot \vec{n}_1)}{\|r_{12}\|^4}
\end{aligned}$$

On voit ainsi que pour que l'on puisse appliquer la formule de Stokes, c'est-à-dire pour que le champ vectoriel $\vec{F}(\vec{r}_{12})$ ait une divergence nulle, il faut et il suffit que $\nabla B(y) \cdot \vec{r}_{12}$ soit nul pour tout y choisi sur A_2. La seule possibilité pour que ceci soit vrai pour plusieurs points x différents est que $\nabla B(y)$ soit identiquement nul sur la surface de l'émetteur, donc que la radiosité de l'émetteur soit constante.

Lorsque $\nabla B(y) = 0$, $\nabla \cdot \vec{F}(\vec{r}_{12}) = 0$, et donc :

$$B(y) \frac{(\vec{r}_{12} \cdot \vec{n}_1)\vec{r}_{12}}{\|r_{12}\|^4} = B(y) \nabla \times \frac{\vec{r}_{12} \times \vec{n}_1}{\|\vec{r}_{12}\|^2}$$

Par là-même, en appliquant le théorème de Stokes,

$$B(x) = E(x) + \frac{\rho}{2\pi} B_0 \oint_{\partial A_2} \frac{\vec{r}_{12} \times \vec{n}_1}{\|\vec{r}_{12}\|^2} \cdot d\vec{\ell}_2$$

On reconnaît dans l'expression ci-dessus un produit mixte $(\vec{r}_{12}, \vec{n}_1, d\vec{\ell}_2)$. D'où l'expression finale (voir équation 4.7):

$$B(x) = E(x) - \vec{n}_1 \cdot \frac{\rho}{2\pi} B_0 \oint_{\partial A_2} \frac{\vec{r}_{12} \times d\vec{\ell}_2}{\|\vec{r}_{12}\|^2}$$

Que pouvons-nous faire lorsque la radiosité n'est pas constante sur l'émetteur, et que nous voulons tout de même retrouver un intégrale de contour ? En tout état de cause, il est toujours possible d'écrire :

$$\vec{F}(\vec{r}_{12}) = \nabla \times \left(B(y) \frac{\vec{r}_{12} \times \vec{n}_1}{\|\vec{r}_{12}\|^2} \right) + \vec{T}$$

où $\vec{T}$ est un terme complémentaire exprimant la différence entre le rotationnel et le champ vectoriel. Plus précisément,

$$\vec{T} = -\nabla B \times \frac{\vec{r}_{12} \times \vec{n}_1}{\|\vec{r}_{12}\|^2}$$

Dans ces conditions, l'expression de la radiosité au point x devient :

$$B(x) = E(x) - \vec{n}_1 \cdot \frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{\vec{r}_{12} \times d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} - \frac{\rho}{2\pi} \int_{A_2} \left(\nabla B(y) \times \frac{\vec{r}_{12} \times \vec{n}_1}{\|\vec{r}_{12}\|^2} \right) \cdot dA_2$$

La radiosité étant vue ici comme une fonction de deux variables, la composante sur $\vec{n}_2$ de son gradient peut être supposée comme nulle : la radiosité ne varie pas dans la direction $\vec{n}_2$. Pour exprimer ceci, on peut introduire un vecteur $\vec{k}$, tel que : $\nabla B(y) = \vec{k}(y) \times \vec{n}_2$. Naturellement, on choisit $\vec{k}$ tel que : $\vec{k} \cdot \vec{n}_2 = 0$. En fait, $\vec{k}(y) = \vec{n}_2 \times \nabla B(y)$.

Ce vecteur $\vec{k}$ simplifie grandement l'expression de la radiosité au point x :

$$B(x) = E(x) - \vec{n}_1 \cdot \frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{\vec{r}_{12} \times d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} - \vec{n}_1 \cdot \frac{\rho}{2\pi} \int_{A_2} \left(\frac{\vec{k}(y) \times \vec{r}_{12}}{\|\vec{r}_{12}\|^2} \right) d\vec{A}_2$$

On a ainsi séparé l'expression de la radiosité en un point donné en deux termes, le premier, une intégrale de contour, et le deuxième une intégrale surfacique, mais différente. L'importance du terme correctif varie suivant la géométrie relative de l'émetteur et du récepteur. Nous verrons au chapitre 10 un exemple d'utilisation de cette formule pour calculer le gradient dû à des émetteurs sur lesquels la radiosité n'est pas constante.

4.9 Prospective

Quel est l'intérêt de ces outils issus de l'analyse des fonctions à plusieurs variables pour les problèmes de calculs d'éclairage ? Quelles possibilités avons nous, de les calculer d'une part, et de pouvoir les employer utilement d'autre part ?

Nous allons voir, au chapitre 5, que la radiosité due à un émetteur constant est une fonction très régulière, et en particulier que les points où elle est concave forment un convexe. Ce résultat n'est qu'une conjecture, mais nous verrons, au chapitre 6, que cette conjecture permet de contrôler la précision avec laquelle nous avons modélisé les interactions entre deux facettes, et donc de contrôler l'erreur commise au cours de nos calculs de radiosité. L'implémentation pratique de ce contrôle de l'erreur dans un algorithme de radiosité hiérarchique fera l'objet de notre chapitre 7.

En raison du caractère plus technique des chapitres concernés, les calculs pratiques des dérivées successives de la radiosité dans des cas particuliers où des valeurs précises sont accessibles directement ont été rejetés plus loin, chapitres 8, 9 et 10.

5.

La conjecture de concavité

NOUS VOUDRIONS pouvoir utiliser les outils de dérivation évoqués au chapitre précédent pour déduire des propriétés de la radiosité qui soient valables sur tout un intervalle, ou sur toute une facette, et non plus seulement en un point.

Naturellement, ceci ne peut se faire dans le cas d'une fonction quelconque. Il faut que la fonction étudiée respecte quelques propriétés simples, soit que ces propriétés soient démontrées, soit que l'on les admette dans un premier temps, comme c'est le cas ici.

Nous allons étudier le comportement de la radiosité dans quelques cas simples, à la fois de façon théorique et de façon expérimentale.

5.1 Travaux antérieurs

Les propriétés de la radiosité sont connues de façon quasi intuitive depuis que la méthode est utilisée dans les calculs d'éclairage (voir [30]).

La première étude précise et formelle de la propriété d'*unimodalité*, c'est-à-dire le fait de n'avoir qu'un seul maximum, est due à Drettakis [14, 11, 12, 10, 9, 13]. Drettakis utilise la propriété d'unimodalité pour reconstruire efficacement la radiosité due à un émetteur convexe sur un plan donné.

5.2 Un cas particulier simple : un émetteur ponctuel

Nous allons étudier de façon complète le Jacobien et le Hessien dans un cas particulier très simple. Ce cas particulier nous donnera néanmoins de précieuses informations sur le comportement qu'on peut attendre de configurations plus complexes.

Considérons donc un émetteur ponctuel, et un récepteur plan infini (voir figure 5.1). On appelle h la distance qui sépare le point émetteur du récepteur.

Alors le vecteur $\vec{r}_{12}$ reliant le récepteur à l'émetteur a pour coordonnées $(-x,-y,h)$. La normale à l'émetteur a pour coordonnées $(0,0,1)$, donc :

$$\cos\theta = \frac{h}{r_{12}}$$

Alors la radiosité sur le récepteur s'exprime comme :

$$B(u,v) = K\frac{\cos\theta}{r_{12}^2} = \frac{Kh}{(h^2+u^2+v^2)^{\frac{3}{2}}}$$

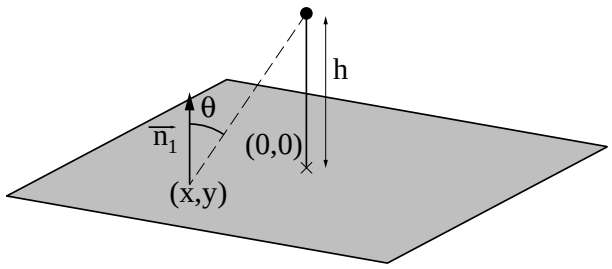


FIG. 5.1 – *Cas d'un point émetteur et d'un plan récepteur.*

où K est une constante comprenant la radiosité du point émetteur, la réflectivité du plan récepteur et un facteur $\frac{1}{\pi}$.

On voit sur la figure 5.2 l'aspect de la radiosité, vue comme une fonction de deux variables u et v.

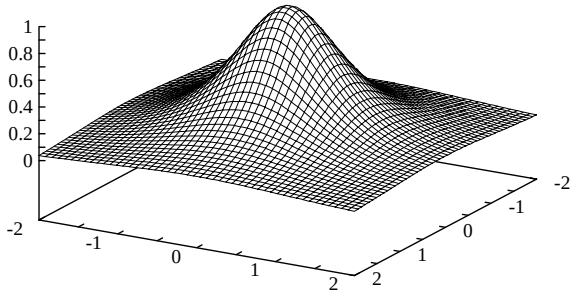


FIG. 5.2 – *Radiosité due à un point émetteur.*

On constate déjà la grande régularité de la fonction, et la présence d'un seul maximum.

Dans un cas simple comme celui-ci, on peut calculer explicitement les dérivées successives :

$$
\begin{aligned}
\frac{\partial B}{\partial u} &= p = -3\frac{Khu}{\left(h^2+u^2+v^2\right)^{\frac{5}{2}}} \\
\frac{\partial B}{\partial v} &= q = -3\frac{Khv}{\left(h^2+u^2+v^2\right)^{\frac{5}{2}}} \\
\frac{\partial^2 B}{\partial u^2} &= r = 3Kh\frac{4u^2-v^2-h^2}{\left(h^2+u^2+v^2\right)^{\frac{7}{2}}} \\
\frac{\partial^2 B}{\partial u \partial v} &= s = 15\frac{Khuv}{\left(h^2+u^2+v^2\right)^{\frac{7}{2}}} \\
\frac{\partial^2 B}{\partial v^2} &= t = 3Kh\frac{4v^2-u^2-h^2}{\left(h^2+u^2+v^2\right)^{\frac{7}{2}}} \\
rt-s^2 &= d = -9K^2h^2\frac{4v^2+4u^2-1}{\left(h^2+u^2+v^2\right)^6}
\end{aligned}
$$

Les représentations graphiques de p, r et $rt-s^2$ peuvent être trouvées sur les figures 5.3, 5.4 et 5.5. Chaque figure porte, outre la représentation de la fonction, la courbe de niveau 0 pour cette fonction, afin que l'on puisse localiser les endroits où elle s'annule. On voit que, sur chaque droite définie par u constant, la dérivée par rapport à u ne s'annule qu'une fois, et la dérivée seconde par rapport à u ne s'annule que deux fois, ce qui fait qu'elle est positive, sauf sur un intervalle borné.

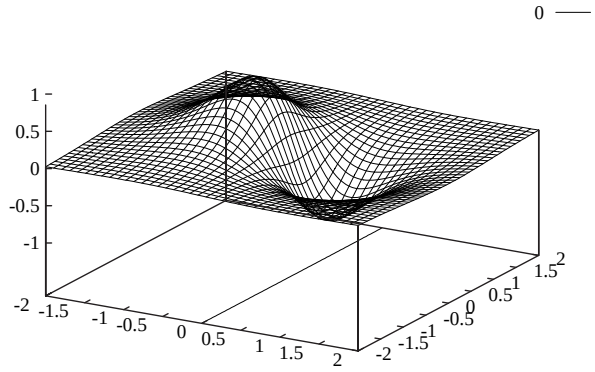


FIG. 5.3 – *La dérivée de la radiosité suivant un vecteur, et la ligne de niveau* 0.

On voit ainsi que d est positif sur une zone convexe, $u^2+v^2 < \frac{h}{4}$, et négatif partout ailleurs, ce qui entraîne que la fonction est concave sur le disque D de centre O et de rayon $\frac{h}{4}$, et traverse son plan tangent partout ailleurs (voir figure 5.6).

Une autre propriété intéressante est l'étude de la fonction de radiosité sur toutes les droites tracées sur le récepteur. Ici, à cause de la symétrie du problème, nous pouvons nous restreindre aux droites définies par u constant.

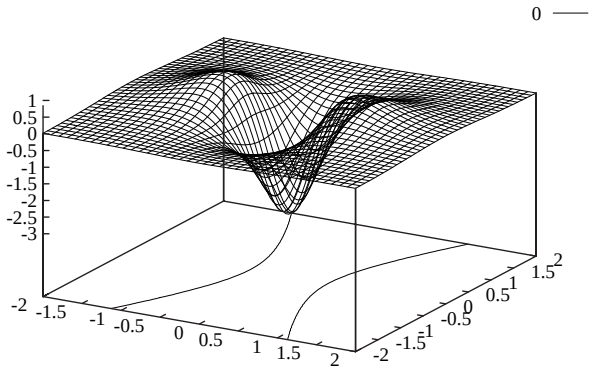


FIG. 5.4 – *La dérivée seconde de la radiosité suivant un vecteur et la ligne de niveau* 0.

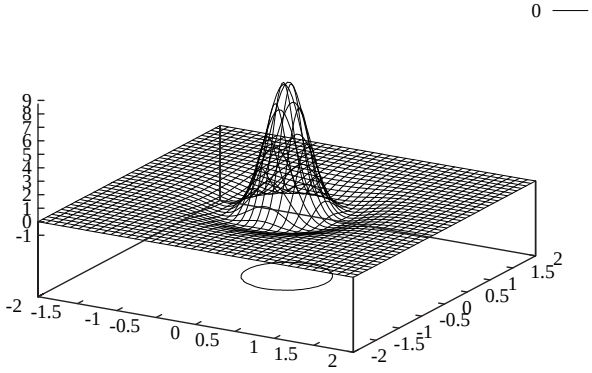


FIG. 5.5 – *La concavité de la radiosité : rt – s^2 et la ligne de niveau* 0.

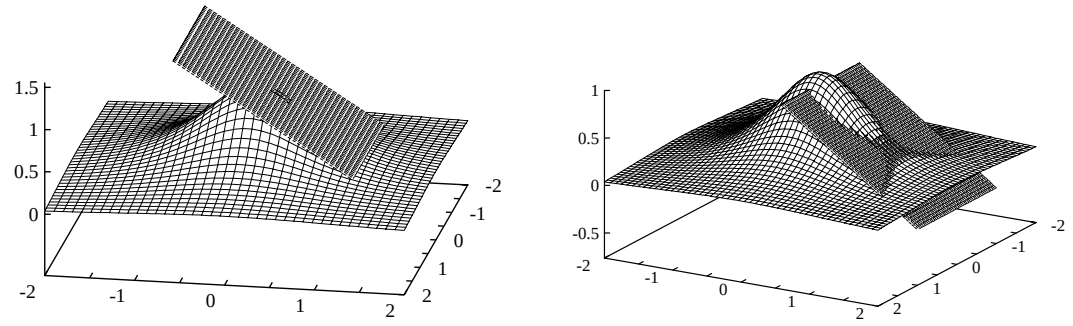


FIG. 5.6 – *La radiosité et ses plans tangents.*

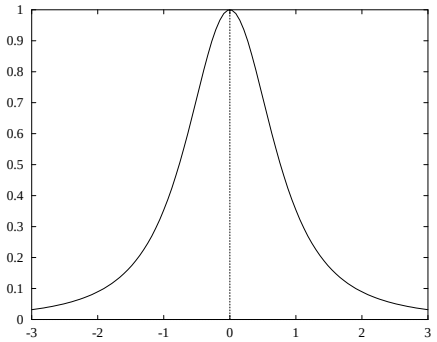
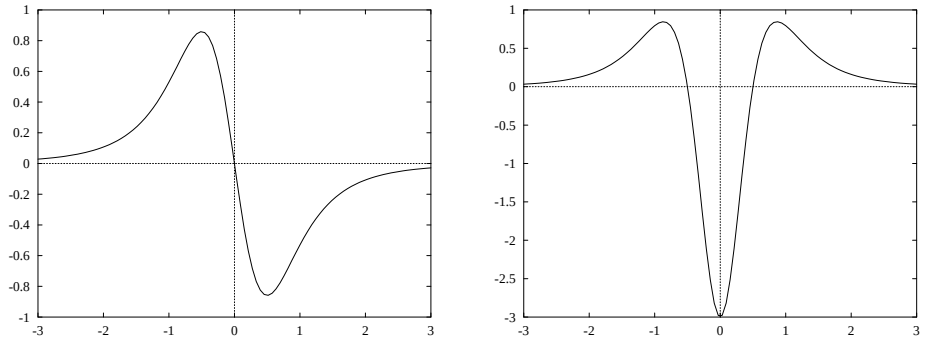


FIG. 5.7 – *La radiosité sur une droite isolée.*

Sur ces droites, la fonction étudiée est une fonction de v, définie par : $f(v) = B(u,v)$ (voir figure 5.7). Sa dérivée $f'(v)$ et sa dérivée seconde $f''(v)$ sont définies par :

$$\begin{aligned} f'(v) &= -3\frac{Khv}{(h^2+u^2+v^2)^{\frac{5}{2}}} \\ f''(v) &= 3Kh\frac{4v^2-u^2-h^2}{(h^2+u^2+v^2)^{\frac{7}{2}}} \end{aligned}$$



a) dérivée b) dérivée seconde

FIG. 5.8 – *Les dérivées de la radiosité par rapport au vecteur directeur de cette droite.*

On voit facilement que $f'(v)$ est du signe de v, et donc ne s'annule qu'en un seul point (voir figure 5.8a). De même, on voit que $f''(v)$ est du signe de $4v^2-u^2-h^2$, et donc est négatif pour

$$v \in \left[-\sqrt{\frac{u^2+h^2}{4}}, \sqrt{\frac{u^2+h^2}{4}}\right]$$

et positif pour v en dehors (voir figure 5.8b).

C'est cette propriété qui est importante : sur une droite donnée, la radiosité considérée comme fonction définie sur cette droite est convexe sauf à l'intérieur d'un intervalle convexe.

Le point émetteur est une bonne approximation de tous les émetteurs, pourvu qu'on les prenne de suffisamment loin. Nous pouvons déjà en déduire que cette propriété devrait se conserver pour tous les émetteurs, à grande distance de l'émetteur.

Par ailleurs, lorsqu'on intègre la radiosité sur un émetteur convexe en situation de visibilité totale, on s'attend également à voir cette propriété conservée, puisqu'il y aura une intégrale de fonctions partageant cette propriété sur un convexe.

On peut aussi étudier la radiosité due non pas à un point émetteur de lumière dans toutes les directions, mais à un élément de surface, orienté et pourvu d'une normale. La radiosité devient :

$$B(u,v) = \frac{Kh(au+bv+ch)}{(h^2+u^2+v^2)^{\frac{3}{2}}}$$

et une étude similaire à celle que nous venons de mener montre que les propriétés de concavité sont bien conservées. Comme la radiosité due à un émetteur convexe est une intégration de telles fonctions, on s'attend à ce que les propriétés de concavité soient bien vérifiées pour un émetteur convexe.

5.3 Le cas d'un disque émetteur

Le cas d'un disque émetteur a été traité de nombreuses fois dans la littérature. La méthode utilisée ici a été en partie empruntée à Moon [30], et avait déjà été partiellement publiée par Drettakis [10].

5.3.1 Préliminaires

Considérons deux éléments de surface, dA_1 et dA_2, situés tous deux sur une sphère de centre O et de rayon R, se faisant face (voir figure 5.9). Comme les normales aux deux points passent par le centre de la sphère, et que trois point sont toujours coplanaires, on peut considérer le plan qui contient O et nos deux points (voir figure 5.9).

On voit que le facteur de forme élémentaire de dA_2 vers dA_1, qui est défini par :

$$dF = \frac{\cos\theta_1 \cos\theta_2}{\pi r^2} dA_2$$

vaut dans ce cas de figure :

$$dF = \frac{dA_2}{4\pi R^2} = \frac{1}{4\pi} d\omega$$

où $d\omega$ est l'élément d'angle solide sous lequel on voit dA_2 depuis O le centre de la sphère. En effet, $\theta_1 = \theta_2$, et $r = 2R\cos\theta_1$. La résultat fondamental est celui-ci : dF est constant pour tous les points de la sphère. Si donc au lieu d'éléments de surface nous considérons une coupole sphérique (voir figure 5.10), de radiosité constante B_0, émettant en direction d'un élément de surface dA_1 situé sur la même sphère, en intégrant nous avons :

$$B_1 = \rho B_0 \frac{1}{4\pi} \Omega$$

où Ω est l'angle solide sous lequel on voit la coupole sphérique depuis le centre de la sphère. Comme la radiosité en un point ne dépend que de l'angle solide sous lequel on

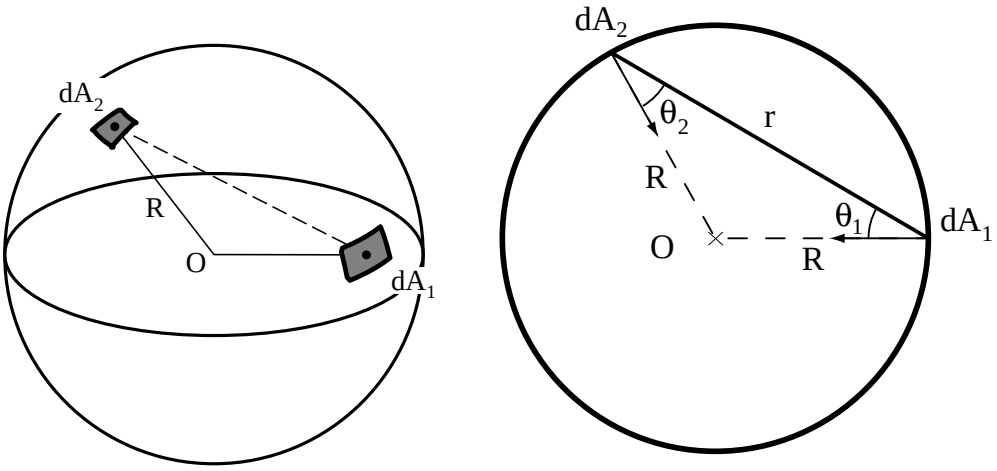


FIG. 5.9 – *Deux éléments de surface d'une sphère en interaction.*

voit l'émetteur, un disque de même contour aura le même effet que la coupole sphérique.

Inversement, si l'on considère un disque émetteur de radiosité constante B_0, et une sphère réceptrice passant par ce disque, la radiosité due au disque sera constante sur tous les points de la sphère. La valeur de la constante dépend bien sûr de la sphère.

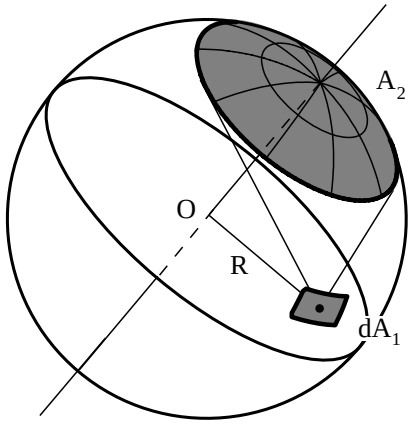


FIG. 5.10 – *Une coupole sphérique illuminant un point de la sphère.*

Notons C le centre du disque, $\vec{n}_2$ la normale au disque, et r le rayon du disque. Toutes les sphères passant par le disque ont leur centre sur la droite portée par $\vec{n}_2$ et passant par C. Prenons S_λ la sphère de centre $O = C + \lambda\vec{n}_2$. Son rayon est $R = \sqrt{r^2 + \lambda^2}$. La surface de la coupole sphérique tendue par le disque est : $A = 2\pi R^2 \frac{R-\lambda}{R}$ (voir figure 5.11). La

radiosité sur la sphère est donc :

$$B = \rho B_0 \frac{R-\lambda}{2R}$$

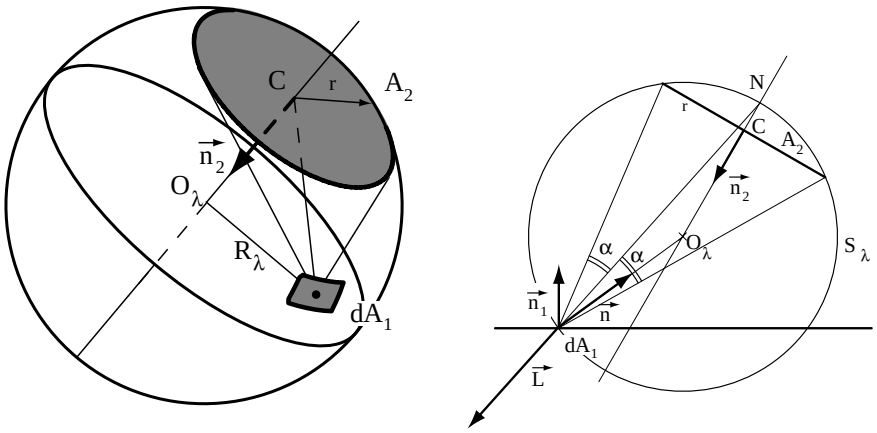


FIG. 5.11 – *Un disque illuminant un point de la sphère.*

On peut voir que la radiosité sur la sphère S_λ est une fonction décroissante de λ.

5.3.2 La radiosité sur le plan

Si maintenant on considère un plan $\mathcal{P}$; ce plan coupe la sphère S_λ suivant un cercle C_λ. En un point M du cercle, la radiosité s'exprime comme :

$$B(M) = -\rho B_0 \int_D \frac{(\vec{r}_{12}\cdot\vec{n}_1)(\vec{r}_{12}\cdot\vec{n}_2)}{r_{12}^2} dA_2 = \vec{n}_1 \cdot \vec{L}$$

Nous connaissons la composante de $\vec{R}$ sur $\vec{n}$, la normale à la sphère au point de contact,

$$\vec{L}\cdot\vec{n} = \rho B_0 \frac{R-\lambda}{2R}$$

Mais ceci est insuffisant pour connaître la valeur de $\vec{L}\cdot\vec{n}_1$. On remarque que l'axe dA_1N est un axe de symétrie pour le cône construit sur le disque émetteur (voir figure 5.11). Le vecteur $\vec{L}$ doit donc se trouver sur cet axe, ce qui nous donne sa direction (N est l'intersection de l'axe du disque avec la sphère S_λ).

Notons $\vec{u}$ le vecteur directeur de $\vec{L}$. On a : $\vec{L} = L\vec{u}$. Donc $\vec{L}\cdot\vec{n} = L\vec{u}\cdot\vec{n}$. Donc :

$$B(M) = (\vec{L}\cdot\vec{n})\frac{\vec{u}\cdot\vec{n}_1}{\vec{u}\cdot\vec{n}} = \rho B_0 \frac{R-\lambda}{2R}\frac{\vec{u}\cdot\vec{n}_1}{\vec{u}\cdot\vec{n}}$$

$\vec{u}$ est porté par $\overrightarrow{MN} = \overrightarrow{MO} + \overrightarrow{ON} = R(\vec{n}-\vec{n}_2)$. Donc :

$$\frac{\vec{u}\cdot\vec{n}_1}{\vec{u}\cdot\vec{n}} = \frac{\vec{n}\cdot\vec{n}_1 - \vec{n}_2\cdot\vec{n}_1}{1-\vec{n}_2\cdot\vec{n}}$$

Et donc la radiosité au point M est :

$$B(M) = \rho B_0 \left(\frac{R-\lambda}{2R} \right) \frac{\vec{n}\cdot\vec{n}_1 - \vec{n}_2\cdot\vec{n}_1}{1-\vec{n}_2\cdot\vec{n}}$$

$\frac{R-\lambda}{2R}$ est une fonction de λ, positive et décroissante. Les sphères S_λ ne coupent le plan considéré que pour λ supérieur à une valeur minimale λ_0.

$\frac{\vec{n}\cdot\vec{n}_1-\vec{n}_2\cdot\vec{n}_1}{1-\vec{n}_2\cdot\vec{n}}$ est une fonction qui dépend à la fois de la sphère coupée et de la position du point M sur le cercle. Seul $\vec{n}_2\cdot\vec{n}$ dépend de la position du point M sur le cercle. On voit que la fraction est maximale lorsque $\vec{n}_2\cdot\vec{n}$ est maximal, donc lorsque le point M est dans le plan qui passe par l'axe du disque. Si l'on note f_λ la valeur maximale de la fraction, on voit que f_λ est une fonction positive décroissante de λ.

La radiosité au point M est donc une fonction positive décroissante de λ. Elle est donc maximale lorsque λ est minimal, donc pour $\lambda = \lambda_0$.

On a ainsi démontré l'unimodalité de la radiosité due à une disque dans le plan. On a par ailleurs la position de ce maximum : sur la sphère passant par le cercle et tangente au plan.

5.4 Le cas d'un carré émetteur

Dans le cas d'un carré d'orientation quelconque, il est impossible de tirer une expression analytique aisément manipulable et dérivable. Comme pour tout polygone émetteur, il existe une expression analytique que l'on peut dériver (voir chapitre 8), et dont on peut même calculer les dérivées successives. En revanche, il est difficile de tirer de cette expression analytique une confirmation – ou une infirmation – de l'unimodalité de la radiosité sur toute droite, encore moins une expression des points où la fonction est concave.

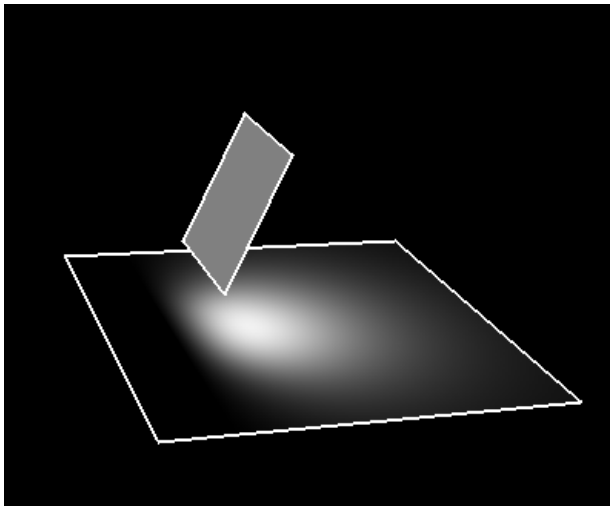


FIG. 5.12 – *Un carré émetteur.*

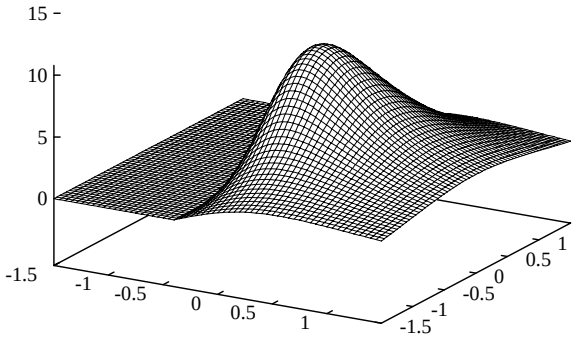


FIG. 5.13 – *Radiosité due à ce carré émetteur.*

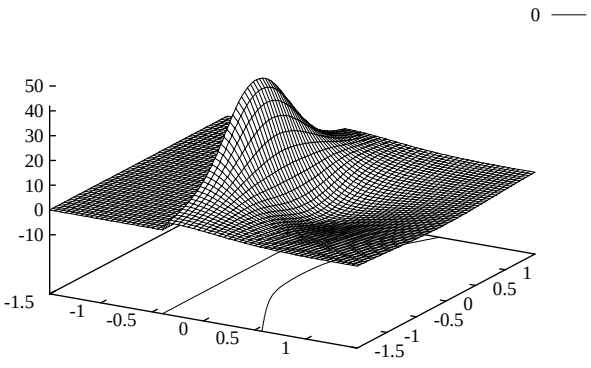


FIG. 5.14 – *Dérivée suivant un vecteur de cette radiosité et la ligne de niveau* 0.

On voit tout de même, sur la figure 5.13, que la radiosité est, en apparence tout du moins, unimodale.

La figure 5.14 représente la dérivée suivant la direction x de la radiosité, avec la courbe de niveau 0 pour cette dérivée. On voit que la dérivée ne s'annule qu'une fois pour chaque droite dans cette direction.

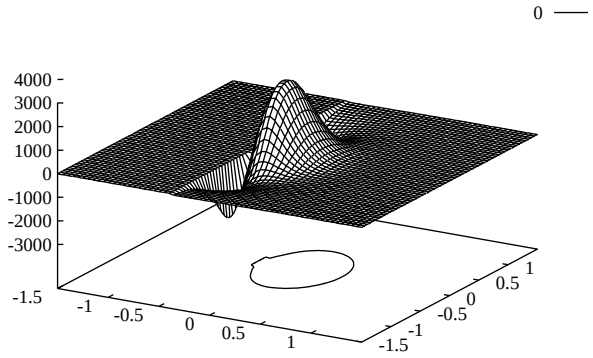


FIG. 5.15 – *Déterminant du Hessien ($rt - s^2$) pour cette radiosité et la ligne de niveau* 0.

La figure 5.15 représente le déterminant du Hessien de la radiosité, avec la courbe de niveau 0. On voit ainsi que ce déterminant est positif sur un convexe, donc que la radiosité est concave sur un convexe, et qu'elle n'est ni concave ni convexe au dehors.

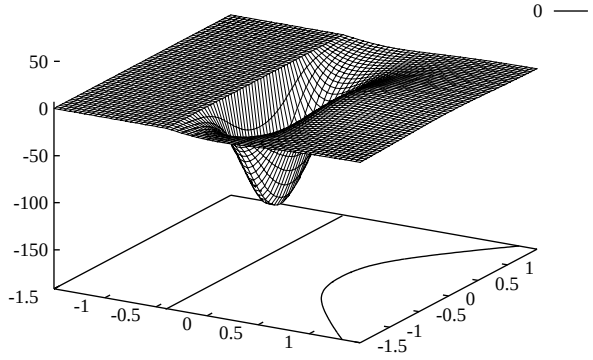


FIG. 5.16 – *Dérivée seconde suivant un vecteur de cette radiosité et la ligne de niveau* 0.

Enfin, la figure 5.16 illustre que la dérivée seconde suivant une vecteur n'est négative que sur un intervalle fini sur chaque droite suivant cette direction.

5.5 Le principe de Curie

Pierre Curie avait conjecturé que lorsqu'un ensemble de causes entraîne un effet, si les causes ont une certaine symétrie (plan de symétrie, axe de symétrie, etc.) alors les effets doivent avoir aussi cette symétrie ; c'est ce qui est généralement résumé par la formule lapidaire : «Les effets ont au moins la symétrie des causes. »

Pour pouvoir appliquer le principe de Curie, il faut que tous les éléments du problème soient symétriques, tant les objets qui interagissent que les interactions qui s'exercent entre eux. On sait qu'il n'existe que quatre interactions possibles à distance entre les objets : la gravitation, l'électromagnétisme, la force forte et la force faible. De ces quatre interactions, les trois premières sont, par nature, symétriques. La force faible peut, dans certains cas, être intrinsèquement asymétrique. Nous ne nous intéressons ici qu'à la propagation de la lumière, qui relève de l'électromagnétisme, et qui donc est symétrique.

Pour pouvoir appliquer le principe de Curie, il faut aussi que la géométrie du problème soit symétrique : il y a d'une part l'émetteur, avec sa géométrie propre, et d'autre part le récepteur et sa géométrie. Pour pouvoir parler de « symétrie du problème », il faut que l'on trouve un plan de symétrie qui conserve à la fois l'émetteur et le récepteur. Nous avons supposé le récepteur comme un plan infini, mais encore faut-il que sa normale soit conservée par le plan de symétrie considéré.

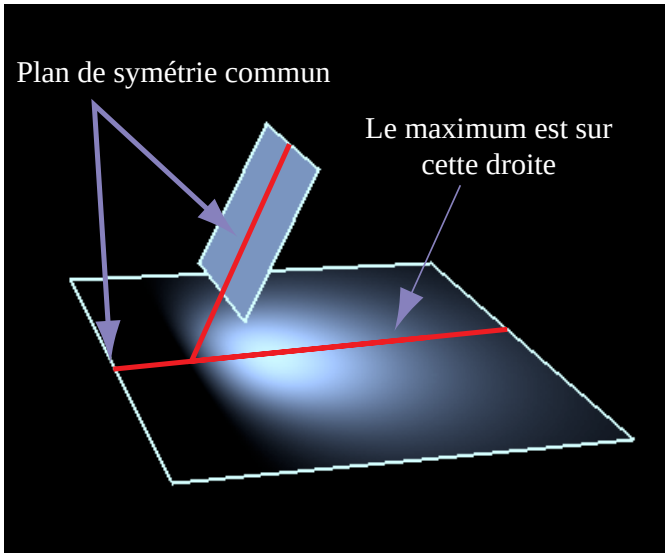


FIG. 5.17 – *Un émetteur possédant un plan de symétrie commun avec le récepteur : le maximum se trouve sur une droite.*

Ainsi, si l'on considère un émetteur ayant un plan de symétrie, il faut pour pouvoir appliquer le principe de Curie que le récepteur soit orthogonal à ce plan de symétrie (voir figure 5.17). Dans ce cas, la radiosité sur l'émetteur aura la même symétrie. C'est à dire qu'elle sera la même pour deux points symétriques l'un de l'autre par rapport à ce plan. Même dans ce cas, on ne sait pas s'il existe un ou plusieurs maximums pour la radiosité, mais on sait que s'il n'en existe qu'un, il sera sur ce plan de symétrie. Et

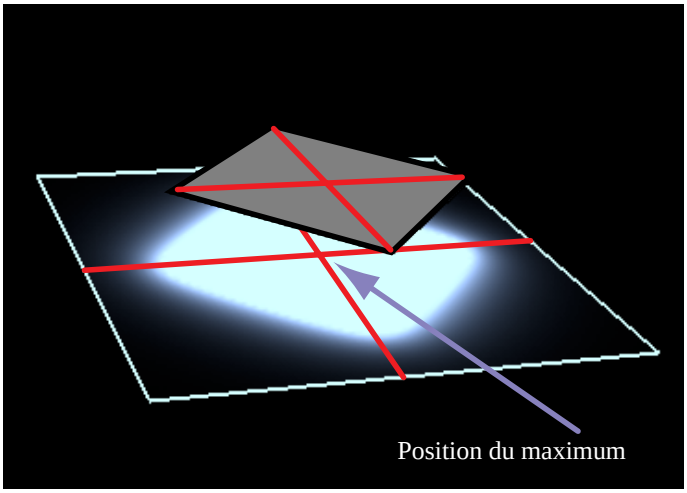


FIG. 5.18 – *Un émetteur possédant deux plans de symétrie commun avec le récepteur : le maximum est localisé.*

s'il en existe plusieurs, il seront symétriques les uns des autres par rapport au plan de symétrie.

Si l'on considère un émetteur ayant deux plans de symétrie (un rectangle, par exemple), il faut pour appliquer le principe de Curie aux deux plans que le récepteur soit orthogonal à ces deux plans. C'est-à-dire qu'il soit parallèle à l'émetteur (voir figure 5.18). Dans ce cas, la radiosité sur le récepteur aura deux plans de symétrie. Ceci ne suffit pas à démontrer l'unicité du maximum, mais permet de le localiser s'il est unique. Notre problème peut être vu comme une extension du principe de Curie : est-ce que le résultat d'un problème convexe sera convexe par nature ?

5.6 La conjecture de concavité

On sait que Drettakis, en 1994 [10] avait conjecturé que la radiosité due à un émetteur convexe ne pouvait avoir qu'un seul maximum. Jusqu'à aujourd'hui, cette conjecture n'a pu être prise en défaut, alors qu'il est très facile d'exhiber un exemple d'émetteur non convexe dont la radiosité n'est pas unimodale : voir figure 5.19.

On a vu sur les exemples précédents que cette conjecture est vérifiée dans des cas simples (disques, carrés).

Au vu des expériences précédentes, on est tenté d'introduire deux conjectures supplémentaires, étendant toutes deux la conjecture d'unimodalité :

> **Conjecture 5.1 :** Sur une droite donnée (appartenant à un récepteur donné), la radiosité due à un émetteur convexe est convexe, sauf sur un seul intervalle borné, où elle est concave.

> **Conjecture 5.2 :** Sur un récepteur plan donné, le Hessien de la radiosité due à un émetteur convexe n'est pas défini, sauf sur un convexe borné, où il est défini-négatif.

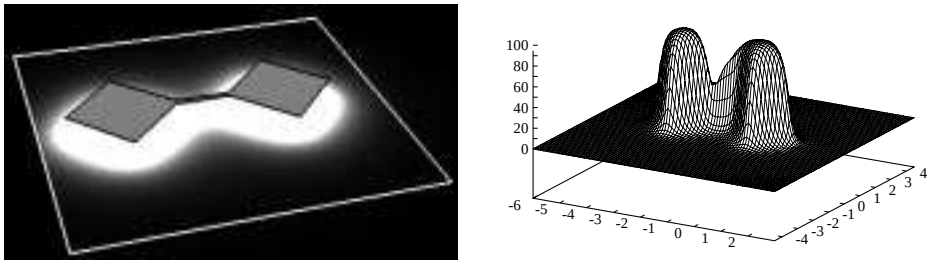


FIG. 5.19 – *Un émetteur qui n'est pas convexe rompt la conjecture.*

On remarque la grande similarité entre ces deux conjectures ; elles ne sont pas équivalentes, mais une méthode utilisée pour démontrer l'une devrait pouvoir être adaptée pour démontrer l'autre. Chacune d'elle a son utilisation pratique, ainsi que nous le verrons dans le chapitre 6.

Par ailleurs, chacune d'elle contient la conjecture d'unimodalité : la conjecture 5.1 implique que sur chaque droite il n'y a qu'un seul maximum (unimodalité sur chaque droite), et donc qu'il n'y a qu'un seul maximum dans le plan.

La conjecture 5.2 implique directement l'unimodalité, car s'il y avait deux maximums, il y aurait entre eux une zone où le Hessien ne serait pas défini négatif, alors qu'il l'est au voisinage de chaque maximum local, donc la zone où le Hessien est défini négatif n'est pas convexe. La conjecture 5.2 n'implique pas l'unimodalité sur chaque droite.

Notons que la conjecture d'unimodalité sur chaque droite implique aussi que si la radiosité admet un extremum sur une droite, cet extremum est nécessairement un maximum : en effet, la radiosité est toujours positive et tend vers zéro lorsque la distance qui sépare le point de l'émetteur tend vers l'infini. Si la radiosité sur une droite avait un minimum, alors entre l'emplacement de ce minimum et plus l'infini on aurait nécessairement un maximum. De même entre ce minimum et moins l'infini. Ce qui fait deux maximums sur la droite.

Par extension, si la radiosité sur un plan admet un extremum, alors cet extremum est un maximum. Sinon, on considère une droite passant par le minimum et on reprend le paragraphe précédent.

5.7 Lorsque la visibilité n'est pas totale

Les conjectures introduites plus haut ne s'appliquent évidemment que dans les cas où la visibilité entre l'émetteur et le récepteur est totale. Lorsqu'il y a des obstacles entre eux, la radiosité n'est plus unimodale, même si les obstacles eux-mêmes sont convexes (voir figure 5.20).

En revanche, si le complémentaire de l'obstacles est convexe, c'est-à-dire que la zone par où passe la lumière est concave, alors il semble bien que la radiosité soit unimodale (voir figures 5.21 et 5.22).

En résumé, si la scène entière est par nature convexe, alors il semble que la radiosité soit unimodale ; cette conjecture est sans doute liée aux deux précédentes ; une méthode

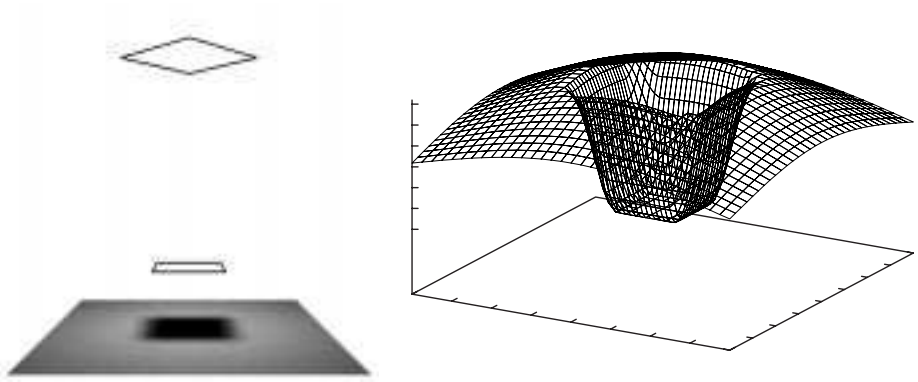


FIG. 5.20 – *Un obstacle, même convexe, remet en cause la conjecture d'unimodalité.*

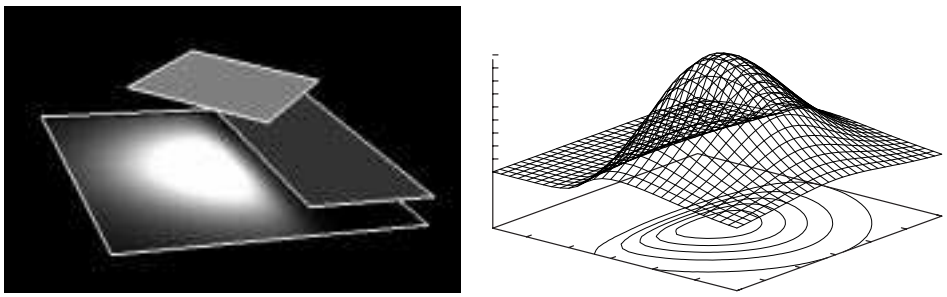


FIG. 5.21 – *Un obstacle qui ne remet pas en cause l'unimodalité.*

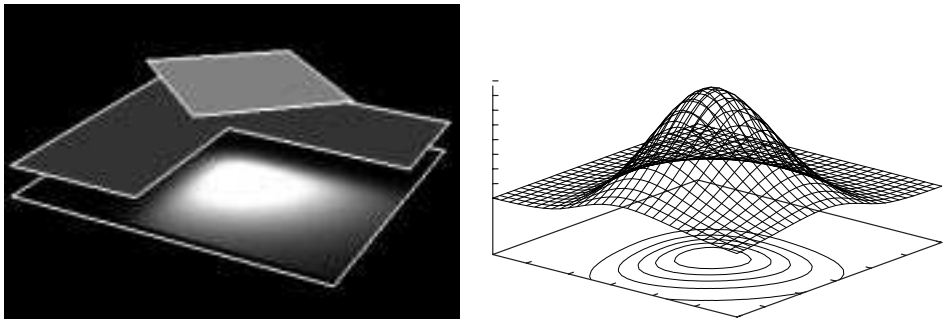


FIG. 5.22 – *Un autre obstacle qui ne remet pas en cause l'unimodalité.*

de démonstration des conjectures 5.1 et 5.2 pourrait sans doute être appliquée pour re-démontrer cette nouvelle conjecture.

6.

Analyse de l'erreur dans les algorithmes de radiosité

LA RÉSOLUTION numérique de la méthode de radiosité ne nous donne qu'une réponse approximative. Il est donc important de pouvoir mesurer la qualité de cette approximation. La mesure de la qualité peut se faire après la fin des calculs, *a posteriori*. Elle sert alors à valider la méthode de radiosité utilisée, à la placer par rapport à d'autres méthodes : par exemple plus rapide, ou plus précise. Elle peut aussi se faire au cours des calculs, *a priori*. La connaissance de l'erreur commise sur la solution calculée permet alors d'orienter les calculs, de les diriger vers les zones où l'imprécision est la plus grande, en s'écartant des zones où la précision est déjà suffisante.

6.1 Travaux antérieurs

Bien que le contrôle de l'erreur au cours de la simulation de l'éclairage global et des calculs de radiosité soit fondamental, et un des points où la méthode de radiosité comporte un avantage précis sur les méthodes de lancer de rayons, la formalisation des calculs d'erreur au cours du raffinement n'a été faite pour la première fois que par Arvo, en 1994 [4] ; Arvo a introduit une taxonomie des erreurs dans un algorithme d'illumination globale, qui est très utile. La première utilisation des calculs de l'erreur au cours du raffinement est due aux travaux de Lischinski [26], en 1994. Ils ont montré comment on pouvait utiliser un encadrement de l'erreur commise sur chaque interaction, à supposer qu'on connaisse cet encadrement, pour calculer l'erreur globale commise sur toute la scène.

L'encadrement, avec exactitude, de l'erreur commise sur chaque interaction, n'a en revanche fait jusqu'ici l'objet d'aucune publication, et les travaux déjà cités de Lischinski utilisent une heuristique, certes efficace, mais qui peut être prise en défaut. Un tel encadrement fait l'objet du chapitre suivant.

6.2 Mesurer l'erreur au cours des calculs de radiosité

Comme nous l'avons vu plus haut, la radiosité est la solution de l'équation intégrale (voir équation 2.11) :

$$B(x) = E(x) + \rho_d(x) \int_{y \in S} B(y) \frac{\cos\theta \cos\theta'}{\pi r^2} V(x,y) dy$$

L'impossibilité de résoudre de telles équations intégrales dans le cas général conduit généralement à une discrétisation du problème. Cette discrétisation nous donne des informations sur la radiosité qui ne peuvent être que limitées : il s'agit des coordonnées de la radiosité dans la base de fonctions choisies pour discrétiser. Si, par exemple, cette base est orthonormée, ce sont les valeurs estimées du produit scalaire de la radiosité avec les fonctions de base choisies. L'information qu'elles apportent est donc limitée à la fonction de base considérée.

C'est à partir de tels échantillons limités que nous devons essayer de reconstruire le comportement de la fonction de radiosité pour deviner la qualité de l'approximation.

Un tel travail est intrinsèquement contradictoire : à partir d'un ensemble d'échantillons, on ne peut pas à la fois déduire une approximation de la radiosité et la qualité de cette approximation. Nous sommes généralement obligés de faire des hypothèses sur le comportement de la radiosité. On peut aussi s'intéresser aux outils de résolution, quantifier leur précision et en déduire celle de la radiosité qu'ils permettent de calculer.

Un protocole de résolution basé sur une estimation de l'erreur commise doit toujours tenir compte de la qualité relative de son approximation. Un calcul de l'erreur, quelle qu'il soit, dépend d'hypothèses diverses, et n'est valide qu'en fonction de ces hypothèses. Il n'est utilisable qu'une fois que les hypothèses nécessaires ont été vérifiées.

6.3 Mesurer l'erreur après les calculs de radiosité

Une autre possibilité de calcul de l'erreur est un calcul *a posteriori*. Il existe différents algorithmes capables de calculer une solution approchée au problème de l'éclairage global. Supposons que nous voulions comparer deux tels algorithmes entre eux ; nous connaissons leurs temps de calcul respectifs, mais comment comparer la qualité des images produites ?

Il ne suffit pas en effet de calculer la différence entre les deux images produites par nos deux algorithmes. Comme ces deux algorithmes ont tous deux produits des solutions qui ne peuvent être qu'approximatives au problème, une différence entre eux n'a pas de sens en tant que telle.

L'idéal serait de pouvoir comparer ces deux algorithmes de synthèse d'image avec une image réelle. On prendrait une photographie d'une scène simple, que l'on pourrait comparer avec les résultats des algorithmes de synthèse. Cette solution n'est encore envisageable que pour des scènes simples, où il est possible de calculer une modélisation précise de la scène, avec notamment les fonctions de réflectance, et la position de la caméra. En effet, il faut qu'un au moins des deux algorithmes à comparer puisse être proche de la solution exacte, faute de quoi la différence entre les deux algorithmes ne fournira pas d'informations exploitables. Se pose également le problème de distinguer entre les erreurs dues à une mauvaise modélisation de la scène, de celles dues à de mauvaises approximations dans les algorithmes de synthèse.

Lorsque cette première solution n'est pas accessible, on peut employer un troisième algorithme de synthèse d'image, qui servira de référence, capable de calculer une solution aussi précise que l'on veut de la scène considérée. On peut alors comparer les deux algorithmes à l'algorithme de référence, et ainsi connaître la différence de précision entre eux. Un bon exemple d'algorithme de référence est un lancer de rayons

distribué (méthode de Monte-Carlo), pour lequel on sait qu'il donne une solution aussi précise que l'on veut – à condition qu'on lui donne suffisamment de temps.

Ce calcul de l'erreur *a posteriori* permet ainsi de valider un algorithme de synthèse d'image, et aussi de le comparer avec les algorithmes déjà existants, à la fois en terme de rapidité et en terme de précision. On peut ainsi mettre en rapport le temps de calcul gagné avec la précision obtenue.

Lorsque l'on compare deux algorithmes de synthèse d'image entre eux, on peut soit comparer les images produites à partir d'un point de vue donné, c'est la comparaison dans l'espace image, soit comparer les résultats obtenus en tous les points de la scène, c'est une comparaison dans l'espace objet. La première solution est plus aisément applicable, puisque de nombreux algorithmes calculent une solution dépendant du point de vue, et non une solution globale.

La comparaison dans l'espace image peut, de façon très simple, commencer par une différence point par point des deux images produites. On a, pour chaque point, la différence entre l'algorithme considéré et l'algorithme de référence. Cette différence ponctuelle peut être utilisée, d'abord dans un but d'analyse, pour permettre de localiser visuellement les faiblesses de l'algorithme considéré, et permettre ainsi de les corriger – sur un plan algorithmique.

Pour quantifier cette différence et permettre ainsi une classification des algorithmes de synthèse d'image, il peut être utile d'avoir une seule grandeur qui quantifie la différence entre les deux images. On peut construire cette grandeur à partir de notre ensemble de valeurs correspondant aux différences ponctuelles, en utilisant une norme vectorielle. Supposons que l'on appelle $\mathcal{E}_i$ l'erreur commise sur le point i par l'algorithme que l'on cherche à quantifier. La norme p des $\mathcal{E}_i$ est :

$$\mathcal{E}_p = \|\mathcal{E}_i\|_p = \left(\sum_i |\mathcal{E}_i|^p \right)^{\frac{1}{p}}$$

Par exemple, la norme 1 ($\|.\|_1$) correspond à la somme des valeurs absolues des erreurs sur chaque point, et la norme ∞ ($\|.\|_\infty$) est égale à la plus grande erreur ponctuelle commise. La norme 2 correspond à une moyenne euclidienne des erreurs ponctuelles.

Ces normes permettent de passer de valeurs ponctuelles multiples à une valeur unique, rendant compte, d'une certaine façon, de toutes les erreurs commises. La norme ∞ ne permet toutefois pas de rendre compte de l'étendue de l'erreur commise : une erreur ponctuelle a peut-être moins d'importance qu'une erreur moindre, mais répétée en de nombreux endroits. À cet égard, la norme 1 permet de tenir compte de l'étendue de l'erreur.

Néanmoins, il faut tenir compte du fait que, souvent, le juge final des images produites est l'œil humain. Or, pour l'œil, toutes les erreurs n'ont pas la même importance suivant leur localisation ; ainsi, une erreur répétée de nombreuses fois en des endroits disséminés (un bruit de fond, par exemple) est plus gênante qu'une erreur, également répétée, mais en des endroits rapprochés. De la même manière, l'œil s'appuie d'abord sur la forme des contours, par exemple les contours des ombres, sur leur régularité. Or il est possible de construire une image telle que la norme 1 de la différence soit inférieure à un seuil donné, et telle que le contour de l'ombre soit déformé.

Les différentes normes p donnent déjà des indications sur la différence entre deux algorithmes, mais ne peuvent complètement mesurer l'erreur visuelle.

6.4 Les différentes sources d'erreur dans un algorithme de radiosité

Pour reprendre une classification introduite par Arvo *et al.* [4], un algorithme de simulation de l'éclairage est sujet à trois principales sources d'erreur :

- **L'erreur de modélisation :** tant les surfaces composant la scène, que leurs échanges d'énergie ou leur émittance ne sont pas forcément correctement pris en compte par le modèle employé.

- **L'erreur de discrétisation :** notre discrétisation du problème ne permet sans doute pas de trouver la solution exacte du problème. Et peut-être n'avons nous pas, même compte tenu de nos hypothèses de discrétisation, choisi la meilleure discrétisation pour approcher la solution.

- **L'erreur de résolution :** notre calcul du système discrétisé lui-même peut n'être qu'approximatif. D'une part, nous ne calculons pas les termes de l'échange avec précision, d'autre part nous ne poussons pas les calculs jusqu'à l'obtention d'une solution exacte.

Idéalement, les trois sources d'erreur devraient être prises en compte et réduites au minimum. Cependant, toutes trois n'interviennent pas aux mêmes étapes de l'algorithme de résolution. Ainsi, l'erreur de modélisation intervient généralement très en amont du processus habituel de calcul. À supposer que nous puissions déterminer qu'une erreur de modélisation à tel endroit aurait des conséquences dramatiques sur les calculs d'éclairage, il serait très difficile pour l'algorithme de résolution d'obtenir des informations complémentaires sur ce point précis. On peut cependant imaginer un algorithme de simulation de l'éclairage disposant de plusieurs niveaux de modélisation des objets de la scène, et capable d'augmenter ou de diminuer le niveau de précision d'un objet en fonction de l'importance de la modélisation de cet objet sur l'erreur globale.

Dans les premiers algorithmes de radiosité, tels ceux de Goral [15] ou de Cohen [6], la discrétisation était effectuée avant la résolution, et il était impossible d'y revenir une fois les calculs commencés. Ce qui obligeait les utilisateurs à effectuer plusieurs étapes modélisation/discrétisation/résolution successives, en tenant compte des erreurs observables sur les premières résolutions pour changer la discrétisation des suivantes.

La plupart des algorithmes de radiosité actuels calculent la discrétisation au vol, au cours du processus de résolution. La fonction chargée d'analyser la solution déjà calculée et d'en déduire les changements à faire, soit dans la discrétisation soit dans les facteurs de forme s'appelle un *oracle,* et elle partage en effet quelques caractéristiques avec les oracles des temps anciens : ses réponses doivent toujours être interprétées, et elles ne sont pas toujours fiables [18].

Un bon oracle de raffinement doit être capable de mesurer l'erreur en cours de résolution, d'indiquer sa nature (erreur de discrétisation ou de résolution) et la meilleure manière de la réduire. Il doit également réduire le nombre de calculs effectués. Ce dernier point est le plus difficile : si l'oracle passe plus de temps à déterminer qu'un calcul est inutile que le calcul lui-même n'en aurait pris, alors c'est l'oracle qui est inutile.

L'oracle doit donc prendre sa décision en se basant au maximum sur les valeurs déjà calculées, et ce travail est très délicat : il s'agit de déterminer le comportement global d'une fonction dont on ne connaît que quelques coordonnées dans une base donnée.

Certaines indications supplémentaires sont éventuellement accessibles, comme par exemple la valeur des échantillons voisins, ou la nature des interactions modélisées (visibilité partielle ou totale). On peut parfois avoir accès à d'autres informations comme les dérivées de radiosité. Il reste à utiliser ces informations de la meilleure manière possible.

6.5 Exemples d'oracles

Naturellement, l'oracle utilisé doit dépendre de l'algorithme utilisé et de ses faiblesses.

Ainsi, l'un des premiers oracles utilisés en radiosité progressive, celui de Cohen, en 1988 [6] effectuait une comparaison des différentes valeurs de radiosité voisines, et raffinait en fonction de la pente de la fonction de radiosité. De la sorte, on trouvait beaucoup d'échantillons dans les zones où la variation de la radiosité était forte, telles que les frontières des ombres.

Nous avons vu plus haut (voir section 2.6.4) les oracles introduits en même temps que la radiosité hiérarchique – bien qu'ils puissent s'adapter aussi à la radiosité progressive.

Le premier, celui d'Hanrahan, en 1990 [16, 17] reposait sur le fait que la méthode utilisée pour approcher les facteurs de forme ne donnait des résultats avec une bonne précision que lorsque les deux facettes concernées étaient petites par rapport à la distance qui les séparait, et en pleine visibilité. Or dans ce cas, le facteur de forme lui-même est petit. On raffine donc lorsque le facteur de forme dépasse une certaine valeur, ε_F. Pour les cas où la visibilité est partielle, on diminue ε_F, pour raffiner plus à ces endroits.

L'introduction de l'importance par Smits *et al.* en 1992 [40] (voir section 2.6.4) reposait sur la même hypothèse, mais quantifiait également l'importance d'une erreur locale sur l'erreur globale.

Pour reprendre la taxonomie d'Arvo *et al.*, on pourrait dire que ces oracles ne s'intéressaient qu'à l'erreur de résolution – même si la solution pour réduire cette erreur consistait à raffiner d'avantage, donc réduisait l'erreur de discrétisation également.

Plus récemment, Lischinski *et al.* [26] utilisent une majoration et une minoration des facteurs de forme pour en déduire un encadrement similaire de la radiosité sur chaque facette. Le traitement consiste ensuite à raffiner les facettes pour lesquelles la différence (majorant moins minorant) dépasse un certain seuil. Cet oracle est le premier exemple de traitement complet de l'erreur. Cependant, il ne s'intéresse qu'à l'erreur de résolution.

Il convient de noter qu'une réduction de l'erreur de résolution entraîne une réduction de l'erreur de discrétisation, la radiosité étant calculée à partir d'une discrétisation plus fine.

6.6 Ce que l'on attend d'un oracle de raffinement

Un bon oracle de raffinement devrait, bien sûr, quantifier à la fois l'erreur de discrétisation et l'erreur de résolution. Il doit également indiquer quelles actions sont les plus à même de réduire ces erreurs, et de combien. Il doit aussi être paramétrable, et permettre ainsi de trouver des solutions plus ou moins précises, suivant le temps qu'on désire y consacrer.

Un oracle doit être précis dans son approximation de l'erreur commise. Lorsqu'en un endroit il surestime l'erreur commise, cela conduit soit à un sur-échantillonnage local, soit à un sous-échantillonnage global : ou bien l'on prend un seuil de raffinement (ε) très petit, et l'algorithme de synthèse d'image raffinera excessivement l'endroit où l'erreur est surestimée, ce qui aboutira à une perte de temps et de ressources système (mémoire) : c'est le sur-échantillonnage local. Si, pour limiter le temps de calcul, on prend un seuil de raffinement élevé, on aura une perte de précision dans les endroits où l'erreur n'est pas surestimée, c'est un sous-échantillonnage global.

De la même manière, si l'oracle sous-estime localement l'erreur commise, cela aboutit soit à un sous-échantillonnage local, soit à un sur-échantillonnage global.

Ce que l'on demande à un oracle de raffinement est donc autant une estimation de l'erreur globale commise que sa localisation précise, et les moyens de la réduire (qui raffiner, et où ?). Cependant, l'importance à accorder à une erreur locale dépend de son importance sur la solution globale, et cette importance dépend de la façon dont cette erreur sera propagée dans la scène.

Par ailleurs, la radiosité calculée comme solution du problème discrétisé n'est pas forcément celle qui sera affichée : même si on a calculé des valeurs constantes par intervalles, on préfère afficher une radiosité variant continûment. La plupart des outils d'affichage accessibles aujourd'hui peuvent effectuer une interpolation linéaire entre des valeurs ponctuelles en utilisant des cartes accélératrices dédiées. Actuellement, très peu de stations de travail donnent accès à une interpolation d'ordre supérieur. Quoi qu'il en soit, la construction de la fonction continue qui sera affichée introduit une différence entre ce qui a été calculé et ce qui est affiché, qui peut selon les cas nous éloigner ou nous rapprocher de la solution de radiosité exacte. L'oracle de raffinement devrait tenir compte de la transformation qui sera effectuée sur la radiosité avant l'affichage, et de son influence sur l'erreur commise, sinon l'algorithme risque à nouveau de sur-échantillonner.

6.7 L'erreur de discrétisation

L'erreur de discrétisation exprime l'impossibilité d'approcher la radiosité avec la discrétisation courante, ou plutôt la différence entre la meilleure solution possible avec la discrétisation courante et la solution réelle.

Lorsqu'on ne connaît pas la solution réelle, on n'a pas accès à une mesure de l'erreur de discrétisation. Ce que l'oracle peut indiquer, en revanche, est la cohérence (ou le manque de cohérence) de la discrétisation actuelle : « compte tenu de ce que nous savons actuellement, et des contraintes suivies par la radiosité, la solution réelle est au moins à cette distance-ci de la solution calculée»

Cette cohérence interne se mesure après une étape de propagation de la radiosité entre toutes les facettes ; on examine alors nos connaissances pour chacune d'elles –

par exemple le majorant absolu et le minorant absolu de la radiosité sur cette facette, et on compare la différence des deux avec notre hypothèse d'une radiosité constante pour la facette.

Si notre hypothèse de discrétisation est différente – par exemple une radiosité variant de façon linéaire, ou des fonctions de base d'ordre supérieur, on peut, au cours du raffinement calculer non seulement une approximation de la radiosité mais aussi une fonction majorante et une fonction minorante. La différence entre ces fonctions majorantes et minorantes donne une mesure de la cohérence interne.

Si l'hypothèse de discrétisation exclut la connaissance de telles fonctions majorantes et minorantes, on peut calculer une approximation ponctuelle des dérivées successives de la radiosité, et comparer une interpolation polynomiale basée sur ces dérivées avec l'hypothèse de discrétisation utilisée.

Mais la cohérence interne ne doit pas se mesurer uniquement sur la discrétisation. Un autre point important est la somme des facteurs de forme qui quittent une facette. Si cette somme est différente de un (1), alors il y a perte ou gain d'énergie au départ de cette facette. Si cette perte est inacceptable – supérieure à un certain seuil – alors il faut affiner notre calcul des facteurs de forme, soit en utilisant une meilleur méthode d'approximation, soit en raffinant la facette.

6.8 L'influence de l'erreur locale sur l'erreur globale

On peut mesurer l'erreur commise sur une interaction donnée, soit avec précision, en utilisant les outils décrits au chapitre suivant, soit approximativement, en utilisant une heuristique basée sur les facteurs de forme ponctuels (voir Lischinski *et al.*, 1994 [26]). L'erreur commise sur une interaction donnée a une influence sur l'erreur commise sur l'ensemble de la scène, mais cette influence dépend de la scène elle-même.

On peut mesurer le lien entre une erreur sur une interaction donnée et l'erreur sur la scène en utilisant un encadrement de la radiosité.

Pour chacun des éléments résultant de la discrétisation en facettes, outre la radiosité estimée B_i, on stocke un majorant de la radiosité, $B_{sup,i}$, et un minorant, $B_{inf,i}$. Les majorants sont des constantes sur chaque facette, mais la radiosité elle-même ne l'est pas forcément : ce critère de calcul de l'erreur globale est utilisable quelle que soit la modélisation de la radiosité sur la facette, pourvu qu'on sache trouver un majorant et un minorant. Outre une modélisation des interactions $\boldsymbol{M}$, on stocke une matrice majorant ces interactions $\boldsymbol{M}_{sup}$, et une matrice qui les minore : $\boldsymbol{M}_{inf}$. De même que $\boldsymbol{B}$, $\boldsymbol{B}_{sup}$ et $\boldsymbol{B}_{inf}$ sont solution d'une équation de transport :

$$\boldsymbol{B}_{sup} = \boldsymbol{E}_{sup} + \boldsymbol{M}_{sup}\boldsymbol{B}_{sup} \tag{6.1}$$

$$\boldsymbol{B}_{inf} = \boldsymbol{E}_{inf} + \boldsymbol{M}_{inf}\boldsymbol{B}_{inf} \tag{6.2}$$

L'erreur sur la radiosité est alors :

$$\delta\boldsymbol{B} = \boldsymbol{B}_{sup} - \boldsymbol{B}_{inf}$$

Cette erreur dépend de la discrétisation choisie. Une erreur sur la radiosité d'une facette de petite taille a moins d'importance pour nous qu'une erreur sur la radiosité d'une facette de grande taille. De plus on ne peut pas la comparer à la radiosité totale de la scène : ça n'aurait pas de sens physique.

Pour toutes ces raison, on introduit l'erreur sur l'énergie :

$$\mathcal{E} = \sum_i (B_{sup,i} - B_{inf,i}) A_i$$

Cette quantité mesure l'erreur commise sur l'énergie totale de la scène. On peut la comparer à l'énergie totale de la scène, ce qui permet de calculer une erreur relative.

On peut, de la même manière, exprimer l'erreur sur une facette i sous la forme de l'énergie dissipée ou créée sur cette facette par nos erreurs de calcul :

$$\begin{aligned} \mathcal{E}_i &= A_i \delta B_i = A_i \left(B_{sup,i} - B_{inf,i} \right) \\ &= A_i \left(E_{sup,i} - E_{inf,i} + \sum_j M_{sup,i,j} B_{sup,j} - M_{inf,i,j} B_{inf,j} \right) \end{aligned} \tag{6.3}$$

Si l'on s'intéresse à l'erreur commise sur toute la scène $\mathcal{E}$, c'est la somme des erreurs commises sur chaque élément ; on peut exprimer cette erreur comme un produit scalaire de deux vecteurs :

$$\mathcal{E} = \sum_i \mathcal{E}_i = {}^t\boldsymbol{A} \delta \boldsymbol{B}$$

(on note $\boldsymbol{A}$ le vecteur des aires A_i).

Notons que $\delta\boldsymbol{B}$ dépend à la fois de $\boldsymbol{B}_{sup}$, de $\boldsymbol{B}_{inf}$, de $\boldsymbol{M}_{sup}$ et de $\boldsymbol{M}_{inf}$:

$$\delta \boldsymbol{B} = \boldsymbol{B}_{sup} - \boldsymbol{B}_{inf} = \boldsymbol{E}_{sup} + \boldsymbol{M}_{sup} \boldsymbol{B}_{sup} - \boldsymbol{E}_{inf} - \boldsymbol{M}_{inf} \boldsymbol{B}_{inf} \tag{6.4}$$

L'équation 6.4 dépend d'une soustraction, et le lien entre $\boldsymbol{M}_{sup}$ et $\boldsymbol{M}_{inf}$ y a été détruit. Il est difficile d'en déduire le lien entre l'erreur en un point et l'erreur globale $\mathcal{E}$. On peut cependant exprimer l'erreur $\mathcal{E}$ sous la forme d'une somme de termes, tels qu'il soit possible séparer la contribution d'une facette i à l'erreur globale, et même la contribution d'une interaction donnée à cette erreur.

Cette dérivation peut se faire de deux manières, soit directement (section 6.8.1), soit en introduisant le concept d'importance (section 6.8.2). Ces deux dérivations ne conduisent pas à la même expression, mais on va voir que les deux expressions que l'on obtient sont très proches.

6.8.1 Dérivation directe

On peut exprimer $\delta\boldsymbol{B}$ en fonction de $\delta\boldsymbol{E}$ et de $\delta\boldsymbol{M}$:

$$\begin{aligned} \delta \boldsymbol{B} &= \boldsymbol{B}_{sup} - \boldsymbol{B}_{inf} = \boldsymbol{E}_{sup} + \boldsymbol{M}_{sup} \boldsymbol{B}_{sup} - \boldsymbol{E}_{inf} - \boldsymbol{M}_{inf} \boldsymbol{B}_{inf} \\ \boldsymbol{M}_{sup} &= \boldsymbol{M}_{inf} + \delta \boldsymbol{M} \end{aligned}$$

et donc,

$$\delta \boldsymbol{B} = \delta \boldsymbol{E} + \delta \boldsymbol{M} \boldsymbol{B}_{sup} + M_{inf} \delta \boldsymbol{B}$$

Ce qui permet d'exprimer l'erreur sur une facette donnée comme :

$$\mathcal{E}_i = A_i \left(\delta E_i + \sum_j \delta M_{i,j} B_{sup,j} + M_{inf,i,j} \delta B_j \right)$$

Si l'on choisit de singulariser dans cette somme la partie qui est due à l'interaction entre i et j :

$$\mathcal{E}_{i,j} = A_i \left(\delta M_{i,j} B_{sup,j} + M_{inf,i,j} \delta B_j \right) \tag{6.5}$$

On peut utiliser cette quantité $\mathcal{E}_{i,j}$ dans un critère de raffinement progressif : on choisit de raffiner toutes les interactions telles que :

$$\mathcal{E}_{i,j} > \varepsilon$$

Une fois que l'on a pris la décision de raffiner l'interaction entre les facettes i et j, il faut encore savoir si l'on raffine l'émetteur ou le récepteur. On peut remarquer que notre expression de l'erreur sur l'interaction dépend de deux termes, l'un contenant l'imprécision sur le facteur de forme ($\delta M_{i,j}$), et l'autre contenant l'erreur sur la facette émettrice (δB_j). Une méthode évidente est de raffiner l'émetteur si c'est le terme qui porte sur l'imprécision de l'émetteur qui est prépondérant, et le récepteur dans le cas contraire :

> Si $M_{inf,i,j} \delta B_j > \delta M_{i,j} B_{sup,j}$, on raffine l'émetteur (j) ; sinon, on raffine le récepteur, i.

6.8.2 Dérivation en utilisant l'importance

L'idée introduite par Lischinski *et al.* [26] est d'utiliser l'*importance* pour quantifier l'influence de l'erreur sur la scène. On va voir que l'expression de l'erreur que l'on obtient est très proche de celle que nous avons déjà obtenue.

On sait que l'importance (introduite pour la première fois par Smits [40]) $\boldsymbol{Z}$ est la solution de l'équation de transport transposée de l'équation de radiosité :

$$\begin{aligned} \boldsymbol{B} &= \boldsymbol{E} + \boldsymbol{M}\boldsymbol{B} \\ \boldsymbol{Z} &= \boldsymbol{R} + {}^{t}\boldsymbol{M}\boldsymbol{Z} \end{aligned}$$

$\boldsymbol{R}$ étant le vecteur des émissions d'importance, facette par facette (on dit aussi vecteur de réception). $\boldsymbol{Z}$ est alors l'importance de chaque facette. Ce qui correspond à l'influence qu'aura une facette sur la radiosité des facettes qui émettent de l'importance.

Une propriété utile de l'importance est que si l'on a une somme des radiosités sur chaque facette pondérée par un coefficient :

$$S = {}^{t}\boldsymbol{P}\boldsymbol{B} = \sum_i P_i B_i$$

alors cette somme est aussi calculable en utilisant l'importance $\boldsymbol{Z}$ obtenue en prenant $\boldsymbol{P}$ comme vecteur des émissions d'importance :

$$\begin{aligned} \boldsymbol{Z} &= (\boldsymbol{I} - {}^{t}\boldsymbol{M})^{-1}\boldsymbol{P} \\ S &= {}^{t}\boldsymbol{P}\boldsymbol{B} = {}^{t}\boldsymbol{Z}\boldsymbol{E} \end{aligned}$$

En effet (voir Smits [40]),

$$\begin{aligned} S &= {}^{t}\boldsymbol{P}\boldsymbol{B} = {}^{t}\left[(\boldsymbol{I} - {}^{t}\boldsymbol{M})\boldsymbol{Z} \right] \boldsymbol{B} \\ &= {}^{t}\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{B} = {}^{t}\boldsymbol{Z}\boldsymbol{E} \end{aligned}$$

Lischinski [26] propose ainsi d'utiliser l'importance pour faciliter le calcul de l'erreur $\mathcal{E}_i$ (introduite à l'équation 6.3).

$$\mathcal{E} = \sum_i A_i \delta B_i = {}^t\boldsymbol{A}\delta\boldsymbol{B}$$

Par ailleurs,

$$\boldsymbol{M}_{sup}\boldsymbol{B}_{sup} = \boldsymbol{B}_{sup} - \boldsymbol{E}_{sup} = (\boldsymbol{M}_{inf} + \delta\boldsymbol{M})\boldsymbol{B}_{sup}$$

Donc,

$$(\boldsymbol{I} - \boldsymbol{M}_{inf})\boldsymbol{B}_{sup} = \boldsymbol{E}_{sup} + \delta\boldsymbol{M}\boldsymbol{B}_{sup}$$

Si nous calculons l'importance à partir des aires et de $\boldsymbol{M}_{inf}$, nous avons :

$$\begin{aligned} (\boldsymbol{I} - {}^t\boldsymbol{M}_{inf})\boldsymbol{Z} &= \boldsymbol{A} \\ {}^t\boldsymbol{A}\boldsymbol{E}_{inf} &= {}^t\boldsymbol{Z}\boldsymbol{B}_{inf} \end{aligned}$$

Donc :

$$\begin{aligned} \mathcal{E} &= {}^t\boldsymbol{A}\delta\boldsymbol{B} = {}^t\boldsymbol{A}\boldsymbol{B}_{sup} - {}^t\boldsymbol{A}\boldsymbol{B}_{inf} \\ &= {}^t\boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{M}_{inf})\boldsymbol{B}_{sup} - {}^t\boldsymbol{Z}\boldsymbol{E}_{inf} \\ &= {}^t\boldsymbol{Z}\boldsymbol{E}_{sup} + {}^t\boldsymbol{Z}\delta\boldsymbol{M}\boldsymbol{B}_{sup} - {}^t\boldsymbol{Z}\boldsymbol{E}_{inf} \\ &= {}^t\boldsymbol{Z}(\delta\boldsymbol{M}\boldsymbol{B}_{sup} + \delta\boldsymbol{E}) \end{aligned}$$

Ici, $\mathcal{E}$ désigne toujours l'erreur effectuée sur l'énergie présente dans la scène. Ce qui est intéressant, c'est que cette erreur (globale, donc) est présentée comme un produit scalaire de deux vecteurs (${}^t\boldsymbol{Z}$ et $(\delta\boldsymbol{M}\boldsymbol{B}_{sup} + \delta\boldsymbol{E})$). Ce produit scalaire s'exprime comme une somme :

$$\mathcal{E} = \sum_i Z_i \left(\sum_j \delta M_{ij} B_{sup,j} + \delta E_i \right)$$

On peut choisir de singulariser le i-ème élément de cette somme, et de l'appeler «erreur sur i» :

$$\mathcal{E}_i = Z_i \left(\delta E_i + \sum_j \delta M_{ij} B_{sup,j} \right)$$

Et, comme nous l'avons fait à l'équation 6.5, on peut s'intéresser au terme de cette somme qui ne dépend que de i et de j :

$$\mathcal{E}_{i,j} = Z_i \delta M_{ij} B_{sup,j}$$

(cette quantité est différente de la quantité $\mathcal{E}_{i,j}$ introduite à l'équation 6.5). Lischinski [26] préconise d'introduire $\mathcal{E}_{i,j}$ dans un critère de raffinement progressif : on raffine toutes les interactions telles que : $\mathcal{E}_{i,j} > \varepsilon$.

On peut voir que cette expression de l'erreur sur l'interaction obtenue par Lischinski est très proche de l'expression de l'erreur obtenue à l'équation 6.5. En effet, Z_i a été obtenu par une équation de transport à partir d'un vecteur d'émission $\boldsymbol{A}$. Z_i est donc au moins égal à A_i, et donc le terme tenant compte de l'imprécision sur $\boldsymbol{M}$, $Z_i \delta M_{i,j} B_{sup,j}$ est au moins égal à $A_i \delta M_{i,j} B_{sup,j}$. En revanche, le terme dépendant de la variation de

radiosité sur l'émetteur, $M_{inf,i,j}\delta B_j$ a été absorbé dans la version de Lischinski par rapport à la version de l'équation 6.5.

L'argument de Lischinski est que le calcul de l'importance de chaque facette Z_i n'augmente pas beaucoup la complexité des calculs. Notons que si l'on calcule l'importance $\boldsymbol{Z}$ par une résolution itérative de l'équation de transport transposée, alors à la première itération, $Z_i = A_i$: les deux critères se rejoignent, et ils ne divergent éventuellement que pour les itérations suivantes.

Une fois qu'on a décidé de raffiner l'interaction entre i et j, il faut encore choisir entre raffiner l'émetteur, j, et le récepteur i. Lischinski propose de raffiner l'émetteur si $M_{inf,i,j}\delta B_j > \delta M_{i,j} B_{sup,j}$, et le récepteur dans le cas contraire. C'est le critère de choix basé sur l'encadrement de la radiosité que nous avons vu plus haut, page 97.

7.

Contrôle de l'erreur en radiosité à l'aide des dérivées successives

NOUS AVONS vu comment la connaissance de l'erreur commise sur une interaction donnée peut servir de base à un critère de raffinement, et ainsi permettre un contrôle de l'erreur globale commise par l'algorithme de simulation de l'éclairage. Nous allons voir comment calculer de façon précise l'erreur commise sur une intégration donnée, en utilisant les dérivées successives de la radiosité. Nous verrons ensuite comment intégrer ce contrôle de l'erreur dans un algorithme de radiosité hiérarchique, ainsi que les structures de données nécessaires.

7.1 Encadrement de l'erreur commise sur une interaction donnée

Considérons une interaction entre deux objets. Dans un premier temps, nous supposerons ces deux objets seuls dans l'espace ; c'est-à-dire que la visibilité entre eux est totale. Nous verrons à la section suivante (section 7.2) comment nos méthodes doivent être adaptées dans le cadre d'une visibilité partielle entre les deux objets.

Nous ne nous intéressons ici qu'à une direction de l'interaction : l'un des objets est considéré comme un émetteur, l'autre est le récepteur. Nous voulons une approximation précise de l'interaction entre ces deux objets ; en particulier, nous voulons un encadrement de la radiosité sur le récepteur en fonction de la radiosité sur l'émetteur.

Nous connaissons une expression exacte de la radiosité en tout point du récepteur, en fonction de la radiosité sur l'émetteur et de la position respective de l'émetteur et du récepteur :

$$B(x) = E(x) + \frac{\rho}{\pi} \int_{A_2} B(y) \frac{\cos\theta_1 \cos\theta_2}{\|r_{12}\|^2} dA_2$$

Cette quantité est définie sur tout le plan qui porte le récepteur. Lorsque la radiosité de l'émetteur est supposée constante, on a également accès à des valeurs exactes pour les dérivées de la radiosité. Lorsque l'émetteur est par surcroît supposé convexe, la quantité $B(x)$ vérifie les conjectures d'unimodalité et de concavité exposées au chapitre 5.

7.1.1 Le minimum se trouve sur un sommet

La conjecture d'unimodalité implique en particulier que la radiosité ne peut atteindre son minimum que sur la frontière du récepteur, et non sur son intérieur. Dans le

cas d'un récepteur polygonal, le minimum de la radiosité est donc atteint sur une des arêtes du récepteur.

La conjecture d'unimodalité sur toute droite implique alors que le minimum de la radiosité sur un segment de droite (l'arête du récepteur) ne peut être atteint sur l'intérieur de ce segment. Le minimum est donc atteint sur une des extrémités du segment, c'est-à-dire un sommet du récepteur.

Dans le cas d'un récepteur polygonal, le minimum de la radiosité est donc atteint sur l'un des sommets du récepteur.

7.1.2 Localiser le maximum

Sans utiliser les dérivées de la radiosité, il est en revanche impossible de localiser précisément le maximum de la radiosité, et surtout, de savoir s'il se trouve ou non à l'intérieur du récepteur.

En revanche, si l'on connaît en un point x le gradient de la radiosité $\nabla B(x)$, alors une conséquence de la conjecture d'unimodalité sur toutes les droites (voir section 5.6) est que le maximum de radiosité se trouve dans le demi-plan défini par :

$$D_x = \{ y \,|\, xy \cdot \nabla B(x) > 0 \}$$

En effet, s'il en était autrement, sur la droite joignant x au maximum, la radiosité décroîtrait avant de croître à nouveau. On aurait donc un minimum local sur cette droite, et deux maximums locaux au moins, en contradiction avec les hypothèses (voir figure 7.1).

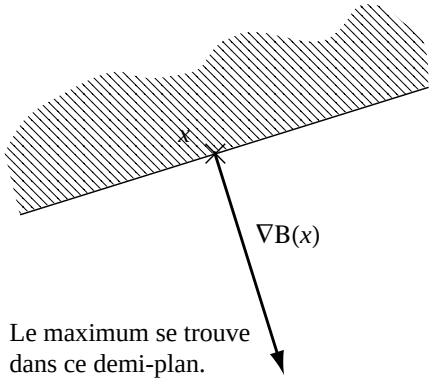


FIG. 7.1 – *Le maximum se situe dans un demi-plan défini par le gradient.*

Lorsqu'on connaît le gradient de radiosité sur tous les sommets du polygone, on a une zone où se situe le maximum : c'est l'intersection de tous les demi-plans D_{x_i} (où les x_i sont les sommets du récepteur).

Si l'intersection des demi-plans et de l'intérieur du récepteur est non-vide, alors le maximum peut se trouver à l'intérieur du récepteur (voir figure 7.2). Si l'intersection des demi-plans et du récepteur est vide, alors le maximum est à l'extérieur du récepteur,

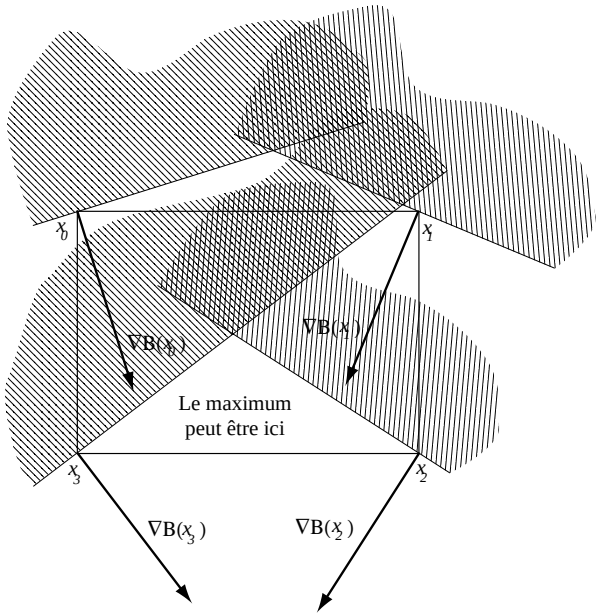


FIG. 7.2 – *Le maximum peut se situer à l'intérieur du polygone.*

et la radiosité atteint donc son maximum sur la frontière – donc sur les arêtes (voir figure 7.3). Qui plus est, en reprenant ce raisonnement, on peut savoir pour chaque arête si le maximum est atteint à l'intérieur de l'arête ou sur un de ses sommets.

Ainsi, sur la figure 7.3, pour trois des arêtes l'intersection entre l'arête et les demi-plans D_{x_i} basés sur ses sommets est vide, et le maximum de la radiosité sur ces arêtes est donc atteint sur un sommet. Pour la quatrième, le maximum de la radiosité est atteint sur l'intérieur de l'arête. Le maximum de la radiosité sur le récepteur est donc le maximum de la radiosité sur les sommets et du maximum de la radiosité sur cette arête.

Comme le récepteur est convexe, pour savoir s'il est contenu dans l'intersection des D_{x_i} il suffit de savoir si ses sommets y sont contenus. Comme il s'agit de demi-plans, pour chaque sommet x_i il suffit de tester si le sommet précédent et le sommet suivant sont inclus.

Le maximum est donc à l'intérieur du récepteur si toutes les conditions suivantes sont vraies, et à l'extérieur dès que l'une d'elle est fausse :

$$\begin{aligned} \forall i,\ \overrightarrow{x_i x_{i+1}} \cdot \nabla B(x_i) &> 0 \\ \forall i,\ \overrightarrow{x_i x_{i-1}} \cdot \nabla B(x_i) &> 0 \end{aligned}$$

7.1.3 Trouver le maximum sur une arête

Si le maximum est atteint à l'extérieur du récepteur, alors le maximum de la radiosité sur le récepteur est nécessairement atteint sur une arête. Pour chaque arête, on peut considérer la radiosité comme une fonction à une variable définie sur l'arête.

Par exemple, si on considère une arête $[AB]$, on a la radiosité comme fonction de

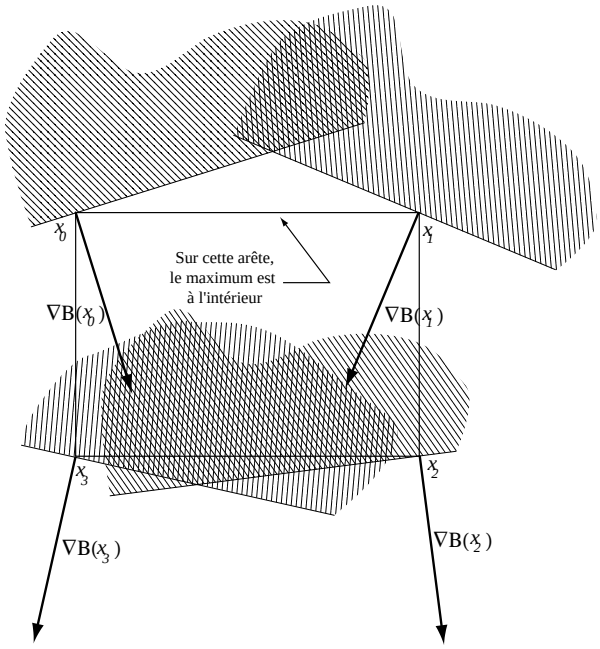


FIG. 7.3 – *Le maximum est atteint sur la frontière.*

l'abscisse sur $[AB]$:

$$f(y) = B(A + y\overrightarrow{AB})$$

La connaissance des dérivées de la radiosité sur les sommets A et B du récepteur nous donne accès aux dérivées de f en 0 et en 1 :

$$\begin{aligned} f'(0) &= \nabla B(A) \cdot \overrightarrow{AB} \\ f'(1) &= \nabla B(B) \cdot \overrightarrow{AB} \\ f''(0) &= {}^t\overrightarrow{AB}\boldsymbol{H}_A\overrightarrow{AB} \\ f''(1) &= {}^t\overrightarrow{AB}\boldsymbol{H}_B\overrightarrow{AB} \end{aligned}$$

À cause de la conjecture d'unimodalité sur toute droite, si $f'(0) \times f'(1) \geq 0$ alors le maximum de la radiosité sur la droite est atteint en dehors du segment $[AB]$, et donc le maximum de la radiosité sur l'arête est atteint sur l'un des sommets.

Inversement, si $f'(0) \times f'(1) < 0$, alors le maximum de la radiosité est atteint sur l'arête elle-même. Si $f''(0) < 0$ et $f''(1) < 0$, alors, à cause de la conjecture de concavité sur les droites (conjecture 5.1), la fonction f est concave sur le segment $[0,1]$. Dans ce cas, elles est majorée par ses tangentes, notamment par les tangentes en 0 et en 1, dont justement on connaît l'expression. On peut donc dire que la radiosité sur l'arête est inférieure à l'ordonnée du point d'intersection des deux tangentes :

$$B_{max} = \frac{f(0)\,f'(1) - f(1)\,f'(0) + f'(0)\,f'(1)}{f'(1) - f'(0)}$$

Si $f'(0) \times f'(1) < 0$, et que $f''(0) \geq 0$ ou $f''(1) \geq 0$, alors on sait que le maximum de la radiosité est atteint sur l'arête, et on sait que les tangentes ne suffisent pas pour obtenir une majoration de la radiosité sur l'arête. On peut alors, soit décider que le facteur de forme ponctuel est majoré par 1, soit calculer une majoration géométrique de la radiosité : on remplace l'émetteur par un autre émetteur, tel que l'on sache que la radiosité due au deuxième émetteur sera toujours supérieure, et tel que l'on sache localiser le maximum de la radiosité due à ce deuxième émetteur, en utilisant le principe de Curie.

On sait que la radiosité en un point dépend uniquement de l'angle solide sous lequel on voit l'émetteur depuis ce point. On prend un plan P_p parallèle au plan récepteur. Dans ce plan, on trace pour les deux sommets de l'arête l'émetteur équivalent à l'émetteur originel ; c'est l'intersection du cône bâti sur le sommet et l'émetteur originel avec le plan P_p. On prend ensuite le vecteur directeur de l'arête, et un vecteur orthogonal également compris dans le plan récepteur. Dans le plan P_p, on construit avec ces deux vecteurs la boîte englobante des deux polygones émetteurs $E(A)$ et $E(B)$ équivalents (voir figure 7.4).

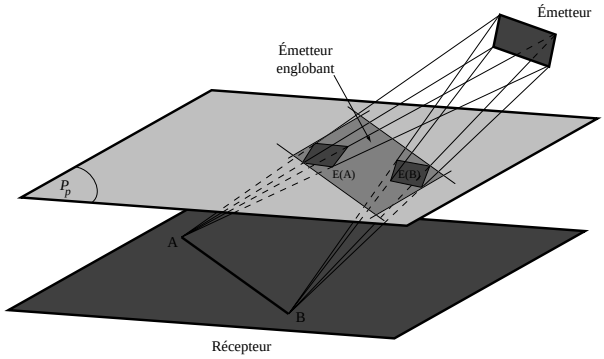


FIG. 7.4 – *Utilisation d'un émetteur englobant pour trouver un majorant sur une arête.*

On a alors un émetteur qui partage sa symétrie avec la droite tracée sur le plan récepteur, donc pour lequel on sait localiser le maximum. On sait aussi que pour tout point situé sur l'arête, entre les deux sommets, la radiosité due à cet émetteur sera supérieure à celle qu'elle serait avec l'émetteur originel, puisque le secteur angulaire solide sous lequel on voit ce nouvel émetteur inclus le secteur angulaire solide sous lequel on voit l'ancien émetteur.

7.1.4 La valeur du maximum sur le récepteur

Si la valeur du gradient de radiosité aux sommets du récepteur indique que le maximum de la radiosité est atteint à l'intérieur du récepteur, on effectue la même classification que dans le cas des arêtes : ou bien la concavité permet de trouver un majorant du maximum, ou bien on a recours à un émetteur majorant, tel que la radiosité qui lui est due majore la radiosité due à l'émetteur originel, et tel que l'on sache localiser le maximum de cette radiosité.

On connaît le Hessien de la radiosité en chacun des sommets du polygone récepteur. On peut savoir à partir du Hessien si la radiosité est concave aux sommets du poly-

gone, en calculant $rt - s^2$. Si la radiosité est concave sur tous les sommets du récepteur ($rt - s^2 < 0$), alors à cause de la conjecture de concavité (conjecture 5.2) on sait que le Hessien sera défini-négatif sur tout l'intérieur du convexe. Donc qu'il sera majoré par tous ses plans tangents, en particulier par les plans tangents aux sommets, dont on connaît l'expression, puisqu'on connaît la valeur du gradient aux sommets.

Nous avons ici un problème de géométrie à résoudre. Mais ce problème (intersection de demi-espaces) est compliqué du fait que nos différentes variables ne sont pas de même nature : nous avons d'un côté la position géométrique des points sommets du récepteur, et d'autre part la radiosité en chacun de ces sommets.

Pour se ramener à un pur problème de géométrie, on transforme chaque couple de valeurs (sommet, radiosité) en un seul point géométrique :

$$V_i = x_i + B(x_i)\vec{n}_1$$

où $\vec{n}_1$ est la normale au plan du récepteur.

On voit immédiatement que la notion «la radiosité en un point est supérieure à...» devient, après cette transformation, «la coordonnée sur $\vec{n}_1$ de ce point est supérieure à...». Si l'on considère que $\vec{n}_1$ est l'axe vertical orienté de notre nouvel espace, dire «la radiosité est inférieure à...» équivaut donc à dire «ce point est en-dessous de...».

L'ordonnée z_M en un point M du plan tangent en un point x_i est, on l'a vu, défini par :

$$z_M = B(x_i) + (M - x_i) \cdot \nabla B(x_i)$$

On a donc :

$$z_M - B(x_i) = (M - x_i) \cdot \nabla B(x_i)$$

Si l'on pose $V_M = M + z_M\vec{n}_1$, on a :

$$(V_M - V_i) \cdot (\vec{n}_1 - \nabla B(x_i)) = 0$$

Dans notre nouvel espace, le plan tangent à la radiosité est donc défini comme le plan passant par V_i, et de normale $\vec{n}_i = \vec{n}_1 - \nabla B(x_i)$.

Pour chaque plan tangent, on peut alors définir le demi-espace négatif, l'ensemble des points qui sont en dessous de ce plan tangent. On sait, puisque la radiosité est concave sur le récepteur, que la radiosité évolue dans l'intersection de ces demi-espaces négatifs. Le maximum de la radiosité est donc majoré par le maximum de cette intersection sur le récepteur. Ce maximum est atteint soit sur l'intérieur du récepteur, soit sur une des arêtes.

Chaque arête définit dans notre nouvel espace un plan, perpendiculaire au plan du récepteur, et donc un demi-espace négatif, tel que l'intérieur du polygone soit l'intersection de tous ces demi-espaces négatifs.

Trouver un majorant de la radiosité en utilisant les plans tangents peut donc se ramener à calculer l'intersection de 8 demi-espaces. Les sommets de cette intersection sont les points d'intersection de trois des plans concernés. Parmi ces points d'intersection, la plupart sont à éliminer car ils se trouvent dans un au moins des demi-espaces positifs.

On calcule donc tous les points d'intersection de trois plans parmi les huit, soit 56 points d'intersection, et on élimine parmi ces points ceux qui se trouvent dans un des demi-espace positif, ce qui fait pour chaque point 5 plans à tester.

En fait, on peut simplifier les calculs dans la mesure où il est inutile de calculer l'intersection de trois plans définis par les arêtes, qui ne peut que se situer en dehors de l'espace intéressant. Notons aussi que l'intersection de deux plans définis par une arête soit correspond à un des sommets du polygone, soit est à éliminer. Si elle correspond à un des sommets du polygone, alors le seul plan tangent à considérer est la plan tangent en ce sommet, puisque tous les autres sont dans son demi-espace positif.

On a donc, d'une part à calculer les intersections de trois plans tangents, soit 4 points, qu'il faut tester avec 5 plans (les quatre plans définis par les arêtes et le plan tangent restant), et d'autre part les intersections de deux plans tangents avec un des plans définis par les arêtes, soit 24 intersections, qu'il faut à chaque fois tester avec les deux plans tangents restants. Une fois qu'on a identifié les sommets V de l'intersection de tous les demi-espaces négatifs, un majorant de la radiosité est donné par la plus grande coordonnée sur $\vec{n}_1$ de ces sommets.

La figure 7.5 résume les étapes du calcul du majorant de la radiosité sur un récepteur où le Hessien est défini-négatif.

Initialisation
pour chaque sommet i du polygone
  $V_i = x_i + B(x_i)\vec{n}_1$
  $\vec{n}_i = \vec{n}_1 - \nabla B(x_i)$
  Ajouter V_i a la liste des sommets.
Sur les aretes
pour chaque arete du polygone
  pour chaque couple de plans tangents (i, j)
    Calculer l'intersection V de i, j et l'arete
    si V est au-dessus d'un des deux autres plans tangents
      Eliminer V
    sinon
      Ajouter V a la liste des sommets.
Les triplets de plans tangents
  pour chaque triplet (i, j, k) de plans tangents
    Calculer l'intersection V des trois plans tangents
    si V est au dessus du quatrieme plan tangent
      Eliminer V
    sinon
      si V correspond a l'interieur du polygone
        Ajouter V a la liste des sommets
      sinon
        Eliminer V
Le majorant
  pour chaque sommet V de la liste
    Calculer la hauteur $h = V \cdot \vec{n}_1$
  Prendre la plus grande valeur h obtenue

FIG. 7.5 – *Majoration avec les plans tangents.*

Lorsque le Hessien n'est pas défini-négatif pour tous les sommets du récepteur, il est impossible de trouver un majorant de la radiosité en utilisant l'intersection des plans tangents. On a alors recours à la même méthode que pour trouver le maximum sur une arête : on prend un plan parallèle au plan du récepteur, sur lequel on trace l'émetteur équivalent à l'émetteur originel pour chacun des sommets du récepteur. On sait qu'alors tout convexe qui contient ces émetteurs équivalents induit sur le récepteur une radio-

sité supérieure à celle de l'émetteur originel. Si on calcule une boîte englobante de ces émetteurs équivalents, on a un émetteur pour lequel on sait localiser, et donc calculer, le maximum de la radiosité sur le récepteur.

7.1.5 Simplification possible

Si le Hessien est défini-négatif sur un triangle, on sait qu'alors la radiosité est minorée sur ce triangle par le plan sécant. Si le récepteur n'est pas un triangle, mais que le Hessien de la radiosité y est défini-négatif, parmi les triangles que l'on peut former avec les sommets du récepteur, il en existe au moins un tel que le plan sécant qu'il définit soit un minorant de la radiosité sur tout le récepteur.

On peut par surcroît choisir parmi tous les plans sécants définis par les triplets de sommets du polygone et qui constituent une minoration de la radiosité celui qui constitue la meilleure minoration, celui qui est tel que la distance entre ce plan sécant et la radiosité des autres sommets soit minimale.

On a alors une minoration de la radiosité par une fonction linéaire. Nous avons vu que la radiosité sera *in fine* affichée comme une fonction linéaire. On peut alors calculer un encadrement de la radiosité sur le récepteur entre d'une part ce plan sécant et d'autre part les plans tangents.

Cet encadrement donne une majoration de la différence entre la radiosité telle qu'elle sera affichée et la valeur exacte de la radiosité sur le récepteur.

En pratique, pour obtenir une majoration, on calcule l'intersection des 8 demi-espaces négatifs, mais pour la dernière étape, au lieu de calculer la hauteur des sommets comme $V \cdot \vec{n}_1$, on calcule la distance entre ces sommets et le plan sécant. Si ce plan sécant est défini par un point V_i et une normale $\vec{n}_s$, la hauteur est $h = \vec{n}_s \cdot \overrightarrow{V_i V}$. On prend comme précédemment la plus grande valeur des h pour tous les sommets.

7.1.6 Résultats

Nous avons implémenté la méthode d'encadrement de l'erreur commise sur une interaction donnée en utilisant le critère d'encadrement de la radiosité décrit ci-dessus. Pour chaque interaction, une fois connue un majorant et un minorant de la radiosité, nous choisissons de raffiner l'interaction si l'erreur sur l'énergie de la facette réceptrice dépasse un certain seuil, c'est-à-dire si :

$$A_i \delta B_i > \varepsilon$$

Le critère heuristique utilisé lorsqu'on ne connaît pas un encadrement de la radiosité sur le récepteur, consiste à calculer la valeur du facteur de forme ponctuel sur les sommets du polygone, que l'on complète par la valeur du facteur de forme ponctuel pris au centre du polygone récepteur. À partir de ces 5 valeurs, on calcule la valeur minimale et la valeur maximale de la radiosité sur le récepteur. Comme nous l'avons vu, la valeur minimale est bien un minorant de la radiosité sur le récepteur, mais la valeur maximale que l'on obtient peut conduire à sous-estimer la radiosité sur le récepteur. Ce critère peut naturellement être pris en défaut dans certains cas particuliers.

Nous avons comparé le raffinement produit par cet algorithme heuristique avec le raffinement produit par les deux algorithmes décrits ici : d'une part l'algorithme qui donne un encadrement exact de la radiosité, avec calcul d'un majorant et d'un minorant, et d'autre part l'algorithme qui, lorsqu'il le peut, c'est-à-dire lorsque la radiosité est

concave, effectue un encadrement par des fonctions linéaires, et qui donc simplifie le maillage produit.

Les tests ont été conduits sur une scène simple, choisie telle que l'algorithme heuristique ne soit pas pris en défaut (voir figure 7.6). Nous avons conduit les calculs pour différentes valeurs de ε.

Pour chaque valeur de ε, nous avons calculé le nombre de facettes produit par les trois algorithmes. Chaque sommet de facette correspond à un calcul de radiosité, donc le nombre de calculs de radiosité effectué par chaque algorithme est approximativement égal au nombre de facettes. Un calcul conjoint de la radiosité, du gradient et du Hessien de la radiosité correspond, en temps de calcul, suivant les ordinateurs, à entre $2,2$ et $2,9$ calculs de radiosité (voir section 8.7). On peut donc calculer, à partir du nombre de facettes, le temps de calcul équivalent de chaque algorithme. Ces valeurs sont regroupées dans le tableau 7.1. On peut voir que le critère de simplification des liens permet un gain en mémoire allant jusqu'à 20 %, et qui augmente à mesure que la précision augmente. On voit aussi que pour ε élevé (1), le critère de raffinement heuristique et notre critère de raffinement complet ne concordent pas, le critère complet décidant de raffiner d'avantage que le critère heuristique.

Les temps de calcul n'ont pas encore été l'objet d'une optimisation poussée : par exemple, si le déterminant du Hessien de la radiosité est positif sur une facette, il est inutile de le recalculer sur les enfants de cette facette, puisqu'il sera positif en vertu de notre conjecture 5.2. Cette optimisation permettrait de réduire le surcoût lié à l'algorithme complet et, dans une moindre mesure, à l'algorithme simplifié.

TAB. 7.1 – *Tests de nos critères de raffinement.*

	Nombre de facettes			Gain dû à la	Surcoût (SGI, standard)	
ε	Heuris.	Complet	Simplifié	simplification	Complet	Simplifié
1	84	120	120	0 %	214 %	214 %
0,1	436	436	420	4 %	120 %	112 %
0,01	1924	1924	1716	11 %	120 %	96 %
0,001	8564	8564	7172	16 %	120 %	84 %
0,0001	43012	43012	34132	21 %	120 %	74 %
0,00001	200386	200386	155708	22 %	120 %	70 %

L'image correspondant à $\varepsilon = 0,0001$ se situe figure 7.6. Comme les trois critères de raffinement coïncident pour tous les points du maillage qui sont communs, avec cette valeur de ε, les trois images sont impossibles à distinguer à l'œil nu, aussi n'en a-t-on fait figurer qu'une seule. Un exemple du maillage produit par les trois algorithmes de raffinement pour deux valeurs de ε, $\varepsilon = 0,01$ et $\varepsilon = 0,0001$ se trouve figure 7.7. On peut voir sur le maillage que les points où le critère de simplification agit sont concentrés dans une zone bien délimitée, celle où le Hessien est défini-négatif.

Pour pouvoir quantifier la différence entre les trois algorithmes, nous avons choisi comme solution de référence la solution produite par l'algorithme complet pour $\varepsilon = 0,00001$. Pour chacune des valeurs de ε, pour les trois algorithmes de raffinement, nous

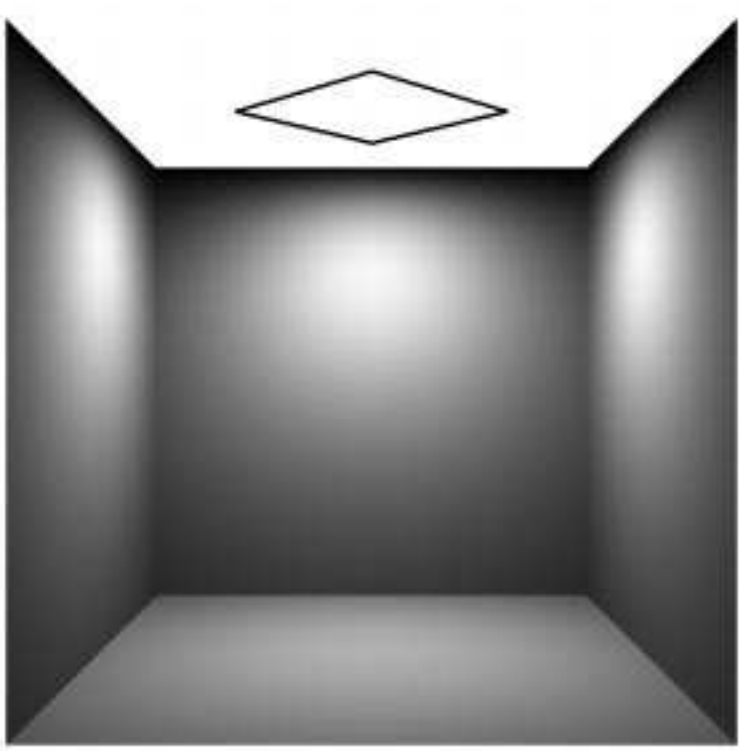
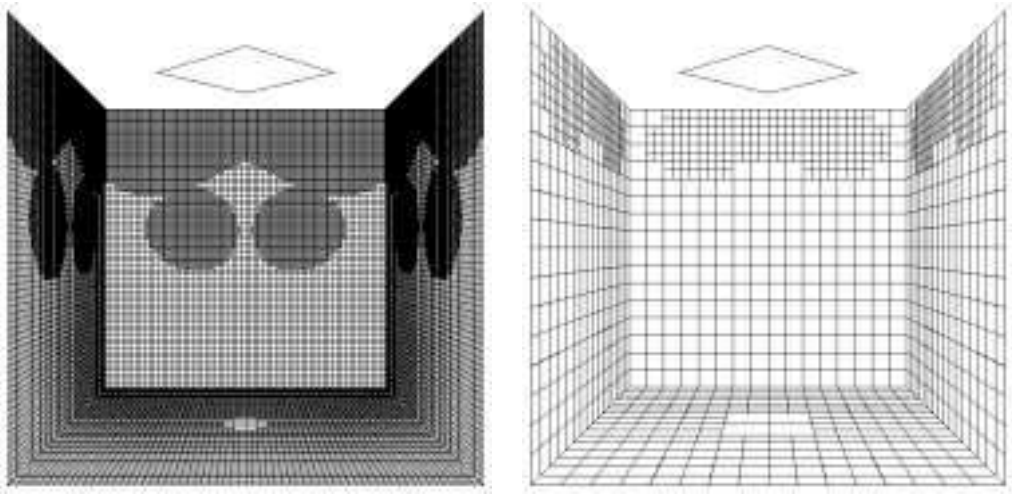
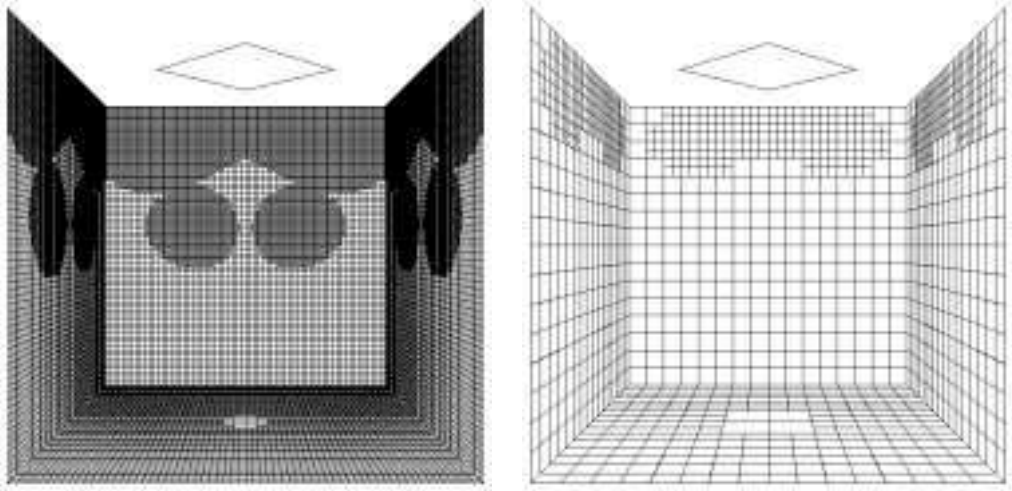


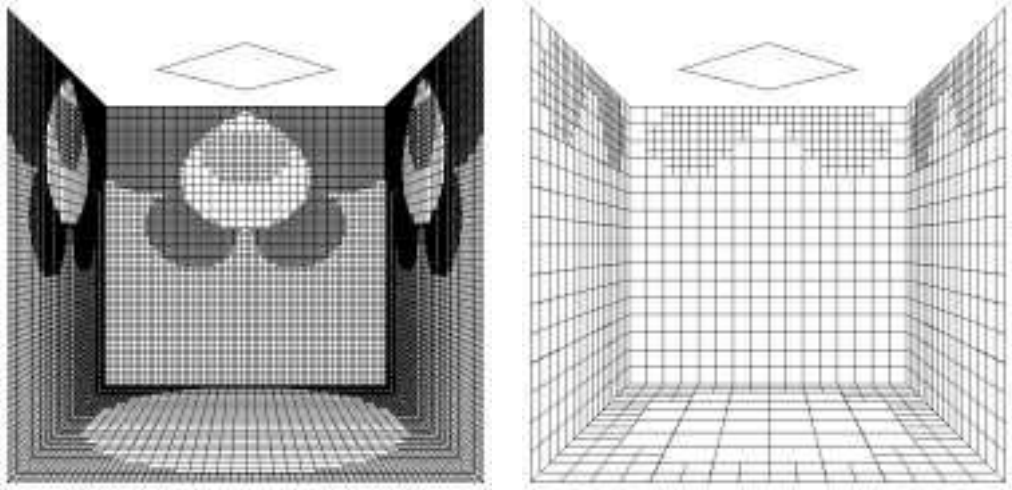
FIG. 7.6 – *La scène de tests pour la visibilité totale,* $\varepsilon = 0,0001$.



Méthode heuristique



Raffinement complet



Raffinement simplifié

FIG. 7.7 – *Le maillage produit par les trois algorithmes,* $\varepsilon = 0,0001$, *et* $\varepsilon = 0,01$.

avons calculé la différence maximale entre la radiosité modélisée par cet algorithme pour cette valeur de ε, et la radiosité de la solution de référence. Cette différence est naturellement une différence entre les interpolations linéaires entre les valeurs aux sommets. Afin d'avoir des valeurs comparables, nous avons pris la différence en termes d'énergie, c'est-à-dire l'intégrale sur la facette de la différence entre la solution de référence et la solution calculée. Il convient de remarquer que cette méthode de calcul attribue – à différence de radiosité égale – une valeur d'autant plus grande que les facettes où il y différence sont larges. Le tableau 7.2 montre cette différence en fonction de la valeurs de ε. Cette différence correspond à l'énergie qui n'est pas modélisée localement.

On voit sur le tableau, d'une part que la différence locale par rapport à la solution de référence est, pour tous les algorithmes, inférieure d'au moins un ordre de grandeur à la valeur de ε. On voit également que l'algorithme simplifié correspond à une modélisation de l'énergie avec la même précision que les algorithmes non simplifiés, puisque la différence n'est pas perceptible.

TAB. 7.2 – *Différence locale de l'énergie pour les algorithmes de raffinement.*

	Heuristique		Simplifié	
ε	Facettes	Erreur	Facettes	Erreur
1,0	84	0,0213689	120	0,213689
0,1	436	0,0015549	420	0,0015549
0,01	1924	0,0015549	1716	0,0015549
0,001	8564	$9{,}71966\times10^{-5}$	7172	$9{,}71966\times10^{-5}$
0,0001	43012	$5{,}06716\times10^{-6}$	34132	$5{,}90167\times10^{-6}$
0,00001	200386	0	200386	$3{,}0568\times10^{-6}$

On peut aussi mesurer l'erreur globale, à partir de l'énergie de la scène. On calcule alors la différence entre l'énergie correspondant à la modélisation de la radiosité que nous avons, et l'énergie de la solution de référence. Cette différence d'énergie est une mesure de l'erreur globale de modélisation effectuée par chaque algorithme. Les résultats de ces mesures sont représentés dans le tableau 7.3. On peut voir que cette erreur globale est très supérieure à la valeur de ε, puisque ε correspond à une majoration seulement locale de l'erreur. On voit également que l'erreur globale due à un algorithme simplifié est supérieure à l'erreur globale due à un algorithme non simplifié, mais que les deux erreur sont du même ordre de grandeur.

En résumé, nous avons présenté deux nouveaux critères de raffinement, permettant tous deux un contrôle de l'erreur commise sur chaque interaction en situation de visibilité totale ; l'un d'eux permet par surcroît une réduction du nombre de facettes utilisées pour modéliser l'interaction, et donc un gain en place mémoire.

Lorsque l'on compare le premier critère à un critère habituel, sur une scène où le critère habituel n'est pas pris en défaut, on voit que le critère avec certitude aboutit au même résultat avec un surcoût de 120 % en temps de calcul. Ce surcoût représente, en quelque sorte, le prix à payer en échange de la certitude.

Notre comparaison, pour avoir un sens, portait sur une scène sur laquelle le critère

TAB. 7.3 – *Mesure de l'énergie totale pour nos critères de raffinement.*

	Heuristique		Complet		Simplifié	
ε	Énergie	Diff.	Énergie	Diff.	Énergie	Diff.
1	27,2370	6,036 %	27,9545	3,949 %	27,9545	3,949 %
0,1	28,7938	1,065 %	28,7938	1,065 %	28,7296	1,286 %
0,01	29,0353	0,235 %	29,0353	0,235 %	28,9987	0,362 %
0,001	29,0897	0,049 %	29,0897	0,049 %	29,0718	0,111 %
0,0001	29,1018	0,008 %	29,1018	0,008 %	29,0932	0,037 %
0,00001	29,1040	0 %	29,1040	0 %	29,0997	0,015 %

heuristique n'était pas pris en défaut. Si l'on comparait les deux critères sur une scène sur laquelle le critère heuristique était pris en défaut, on constaterait une erreur, mesurée en termes d'énergie locale, pouvant aller jusqu'à 100 % pour le critère heuristique, et un accroissement du temps de calcul pour le critère complet par rapport au critère heuristique. Il convient ici de remarquer que parmi les scènes qui peuvent induire en erreur le critère heuristique, il en existe – par exemple lorsqu'un obstacle n'est pas détecté lors du premier raffinement – pour lesquelles réduire le seuil de raffinement ne permettrait pas de réduire l'erreur commise. Inversement, notre critère complet permet toujours de trouver un encadrement exact de la radiosité sur la facette réceptrice, et donc une réduction du seuil de raffinement se traduira toujours par une réduction de l'erreur commise.

Lorsque l'on compare le critère avec réduction des liens au critère habituel, on constate qu'ils aboutissent à des modélisations différentes de la radiosité sur la facette, mais proches cependant en tous points. Ce critère avec réduction des liens permet par ailleurs une réduction de la place mémoire utilisée de l'ordre de 20 %, et augmentant avec la précision.

7.2 Le cas de la visibilité partielle

7.2.1 Maintenir un encadrement de la radiosité

Nous avons vu comment obtenir un encadrement de la radiosité dans le cas de deux objets convexes en interaction, seuls. Lorsqu'un obstacle est présent entre eux, la solution est plus complexe, puisque les conjectures de concavité sur lesquelles nous avons basé notre encadrement ne sont plus accessibles : nous avons vu au chapitre 5 que même un unique obstacle convexe pouvait mettre en défaut la conjecture d'unimodalité sur les droites.

La seule solution pour maintenir un encadrement de la radiosité lorsqu'il y a un obstacle consiste à calculer, d'une part un convexe situé sur l'émetteur, inclus dans l'émetteur, et tel que tous ses points soient en situation de visibilité totale avec le récepteur, que l'on appelle l'émetteur *minimal*, et d'autre part un convexe situé sur l'émetteur, inclus dans l'émetteur, et tel qu'il contienne tous les points de l'émetteur qui sont en situation de visibilité avec au moins un point du récepteur. Ce deuxième convexe est appelé émetteur *maximal*.

On voit que la radiosité due à l'émetteur minimal en situation de visibilité totale constitue une minoration de la radiosité due à l'émetteur en situation de visibilité partielle, et que la radiosité due à l'émetteur maximal en situation de visibilité partielle constitue une majoration.

Comme l'émetteur minimal et l'émetteur maximal sont des convexes en situation de visibilité totale, nous pouvons appliquer les résultats du chapitre précédent pour calculer, d'une part une minoration de la radiosité de l'émetteur minimal, et d'autre part une majoration de la radiosité de l'émetteur maximal.

Le calcul de l'émetteur minimal et de l'émetteur maximal se fait à partir des arêtes des obstacles présents entre l'émetteur et le récepteur. Pour chaque arête prise isolément, on détermine un plan minimal, tel que tous les points situés dans le demi-espace positif de ce plan soient en situation de visibilité totale avec le récepteur, et un plan maximal, tel que tous les points situés dans le demi-espace négatif de ce plan soient en situation d'invisibilité totale avec le récepteur. Le plan maximal est donc celui qui cache le plus le récepteur, et le plan minimal celui qui le cache le moins.

Ce plan minimal et le plan maximal se calculent, à partir de l'arête de l'obstacle et des sommets du récepteur convexe. On prend pour chaque sommet le plan défini par ce sommet et l'arête. Chaque plan possède une normale, et l'on peut classer ces normales en fonction de l'angle orienté qu'elles font avec la normale au plan du récepteur, le trièdre étant orienté par l'arête de l'obstacle. Le plan qui correspond à l'angle le plus petit (généralement négatif) est le plan minimal, et le plan qui correspond à l'angle le plus grand (généralement positif) est le plan maximal.

Lorsque l'obstacle est un unique convexe, on procède de façon incrémentale sur les arêtes de l'obstacle pour trouver l'émetteur minimal :

- si l'émetteur est totalement inclus dans le demi-espace positif, on est en situation de visibilité totale, donc l'émetteur minimal est égal à l'émetteur ;

- si l'émetteur est totalement inclus dans le demi-espace négatif, cette arête n'est pas significative, et on la laisse de côté ;

- si cette arête coupe l'émetteur, on prend la portion de l'émetteur qui est incluse dans le demi-espace positif (voir figure 7.8). Cette portion est un candidat acceptable pour l'émetteur minimal, puisqu'elle est convexe et toujours visible de tous les points du récepteur. On a donc éventuellement plusieurs candidats pour l'émetteur minimal – au maximum quatre. Tous sont acceptables, mais il est préférable de choisir celui qui donnera la minoration la plus élevée, permettant ainsi le plus petit encadrement de la radiosité sur le récepteur. Si l'on ne veut pas calculer une minoration pour chaque candidat émetteur minimal, on peut se contenter de prendre le candidat dont l'aire est la plus grande. Cette heuristique permet de trouver plus rapidement un émetteur minimal dont on peut espérer qu'il donnera une bonne minoration, même si cette minoration ne sera pas nécessairement la plus élevée possible. Si la minoration obtenue n'est pas la meilleure, le résultat sera un raffinement plus grand. Nous avons ici un choix entre deux critères, l'un donnant le meilleur maillage possible, donc économe en mémoire, au prix d'un temps de calcul supplémentaire, et l'autre donnant un maillage qui sera peut-être trop raffiné, et donc plus coûteux en mémoire, mais économisant le temps de calcul.

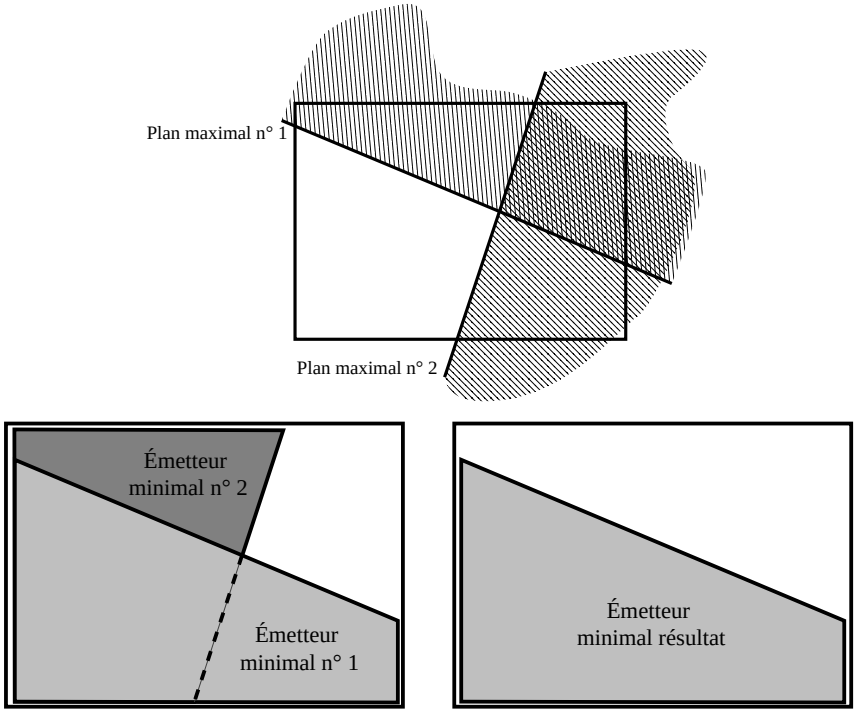


FIG. 7.8 – *L'émetteur minimal est l'un des émetteurs minimaux calculés.*

On procède de la même façon pour trouver l'émetteur maximal :

– si l'émetteur est totalement inclus dans le demi-espace positif, l'émetteur maximal est égal à l'émetteur ;

– si l'émetteur est totalement inclus dans le demi-espace négatif, cette arête n'est pas significative ;

– si le plan minimal coupe l'émetteur, on prend l'enveloppe convexe de l'ancien émetteur maximal et de l'intersection du demi-espace positif et de l'émetteur. On remarque que cette enveloppe convexe est égale à l'émetteur coupé par une arête, donc que le seul travail à accomplir pour calculer l'émetteur maximal est de trouver cette arête ; de plus, cette arête dépend uniquement de l'ancienne arête et de l'arête créée par le plan maximal courant (voir figure 7.9).

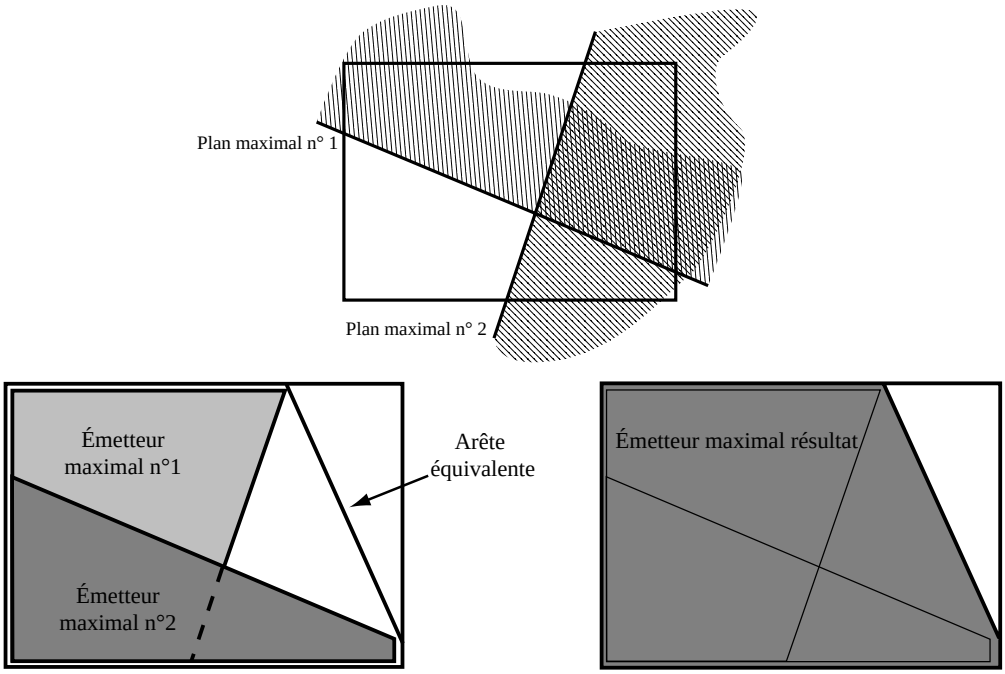


FIG. 7.9 – *L'émetteur maximal se construit comme une enveloppe convexe.*

La figure 7.10 montre un exemple d'émetteur minimal et d'émetteur maximal. Le carré rouge sur le sol est la facette pour laquelle on a calculé la visibilité. La zone rouge clair représente l'émetteur minimal, visible de tous les points de la facette réceptrice. La zone rouge foncée représente l'émetteur maximal, enveloppe convexe de tous les points de l'émetteur visibles d'au moins un point du récepteur. On a rajouté le contour de la partie de l'émetteur qui est visible d'un des sommets de la facette réceptrice. On peut voir que ce contour est bien inclus dans l'émetteur maximal, et contient bien l'émetteur minimal. Le cartouche sur le côté représente les mêmes informations, vues de dessus.

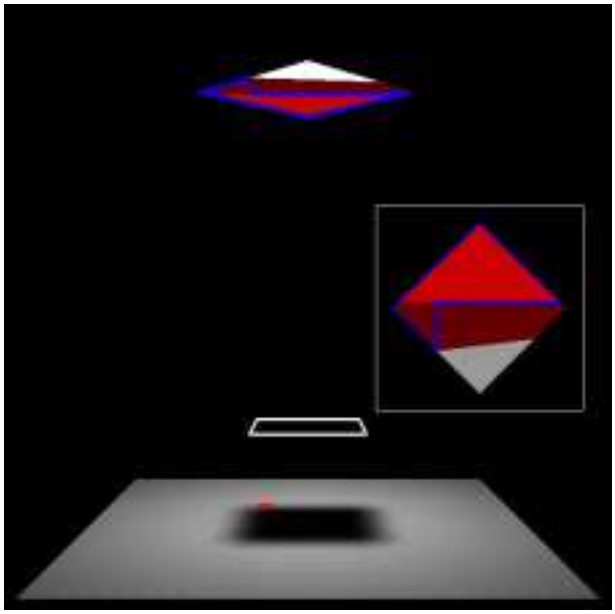


FIG. 7.10 – *Exemple d'émetteur minimal et maximal, avec un des émetteurs exacts.*

7.2.2 Lorsqu'il y a plusieurs obstacles

Lorsqu'il y a plusieurs obstacles, on procède de manière similaire, en cherchant toujours à conserver pour les émetteurs maximal et minimal leurs propriétés essentielles, à savoir qu'ils sont convexes et qu'ils encadrent l'émetteur réel pour tous les points du récepteur.

Pour l'émetteur maximal, on commence par calculer pour chaque obstacle l'union des demi-espaces minimaux : c'est l'ensemble des points de l'émetteur qui sont visible d'au moins un des points du récepteur par rapport à cet obstacle. On prend ensuite l'intersection de ces ensembles : c'est l'ensemble des points de l'émetteur qui sont visibles d'au moins un des points du récepteur compte tenu de tous les obstacles. Enfin, on prend l'enveloppe convexe de cet ensemble : c'est l'émetteur maximal.

Si la complexité de la scène et des obstacles entre l'émetteur et le récepteur rend prohibitif ce calcul précis de l'émetteur maximal, on peut avoir recours à d'autres méthodes, qui donnent un émetteur maximal plus grand, donc un encadrement plus large, mais qui permettent d'aboutir au résultat plus rapidement. Par exemple considérer que l'émetteur maximal est égal à l'émetteur original.

Le calcul pratique de l'émetteur maximal peut se faire sans calculer explicitement d'unions ou d'intersections de polygones : pour le premier obstacle, on sait calculer pour chaque arête l'intersection de l'émetteur et du demi-espace minimal. Ce qui nous donne une liste de polygones convexes, dont l'union représente l'ensemble des points de l'émetteur qui sont visibles d'un point du récepteur. Pour les obstacles suivants, on procède de la même manière, mais on calcule l'intersection du demi-espace minimal avec chacun des polygones convexes présents dans la liste, augmentant ainsi cette même liste. À la fin, on prend l'ensemble des sommets des polygones convexes pré-

sents dans la liste. L'émetteur maximal est égal à l'enveloppe convexe de ces sommets. L'algorithme de calcul de l'émetteur maximal est résumé figure 7.11.

```
initialiser la liste ℓ des emetteurs
ajouter l'emetteur originel       a la liste ℓ
pour chaque obstacle potentiel
    pour chaque emetteur de la liste ℓ
        retirer l'emetteur de la liste ℓ
        pour chaque arete de l'obstacle
            calculer l'intersection de l'emetteur et du demi-plan minimal
            si cette intersection est non vide
                ajouter le resultat de l'intersection a la liste ℓ
initialiser la liste s des sommets
pour chaque emetteur de la liste ℓ
    ajouter ses sommets a la liste s
calculer l'enveloppe convexe des sommets de la liste s
```

FIG. 7.11 – *Calcul de l'émetteur maximal en présence de plusieurs obstacles.*

Pour l'émetteur minimal, on peut soit prendre un convexe qui englobe l'ensemble des obstacles, tel une boîte englobante ou une enveloppe convexe, ou bien calculer un émetteur minimal de la même manière que l'émetteur maximal : pour le premier obstacle, on calcule pour chaque arête l'intersection de l'émetteur avec le demi-espace maximal. Ce qui nous donne une liste de polygones convexes, chacun d'eux étant un candidat valable pour être émetteur minimal. Pour les obstacles suivants, on calcule pour chaque arête de l'obstacle l'intersection du demi-espace maximal avec les émetteurs présents dans la liste, le résultat étant ajouté à la liste elle-même. Lorsque tous les obstacles ont été parcourus, la liste contient un ensemble de candidats valables au titre d'émetteur minimal. Il reste à choisir parmi les différents candidats. Ceci peut se faire, soit en calculant pour chacun d'eux la minoration qu'il induit sur le récepteur, et en prenant celui qui induit la minoration la plus grande, soit en prenant le candidat dont l'aire est la plus élevée. Le pseudo-code de calcul de l'émetteur minimal est résumé figure 7.12.

```
initialiser la liste ℓ des emetteurs
ajouter l'emetteur originel       a la liste ℓ
pour chaque obstacle potentiel
    pour chaque emetteur de la liste ℓ
        retirer l'emetteur de la liste ℓ
        pour chaque arete de l'obstacle
            calculer l'intersection de l'emetteur et du demi-plan maximal
            si cette intersection est non vide
                ajouter le resultat de l'intersection a la liste ℓ
selectionner l'emetteur de la liste ℓ dont l'aire est la plus grande.
```

FIG. 7.12 – *Calcul de l'émetteur minimal en présence de plusieurs obstacles.*

7.2.3 Résultats

Nous avons implémenté le calcul de l'émetteur minimal et maximal, et regroupé ce code avec le calcul d'un majorant et d'un minorant de la radiosité pour obtenir un algorithme de radiosité hiérarchique avec un critère de raffinement basé sur un contrôle de l'erreur, y compris en présence d'obstacles.

La scène choisie pour les tests est une scène simple, composée d'un seul émetteur, d'une seul récepteur et d'un obstacle qui les sépare. Nous avons comparé nos algorithmes (complet et simplifié) avec le même algorithme heuristique : calcul du facteur de forme ponctuel en cinq points, et l'erreur sur cette interaction est prise égale à la différence entre le plus grand des facteurs de forme ponctuels et le plus petit des facteurs de forme ponctuels, multipliée par la radiosité de l'émetteur et l'aire du récepteur.

La figure 7.13 visualise la scène de test employée, et les figures 7.14, 7.15 et 7.16 montrent le maillage produit par les trois algorithmes pour différentes valeurs de ε.

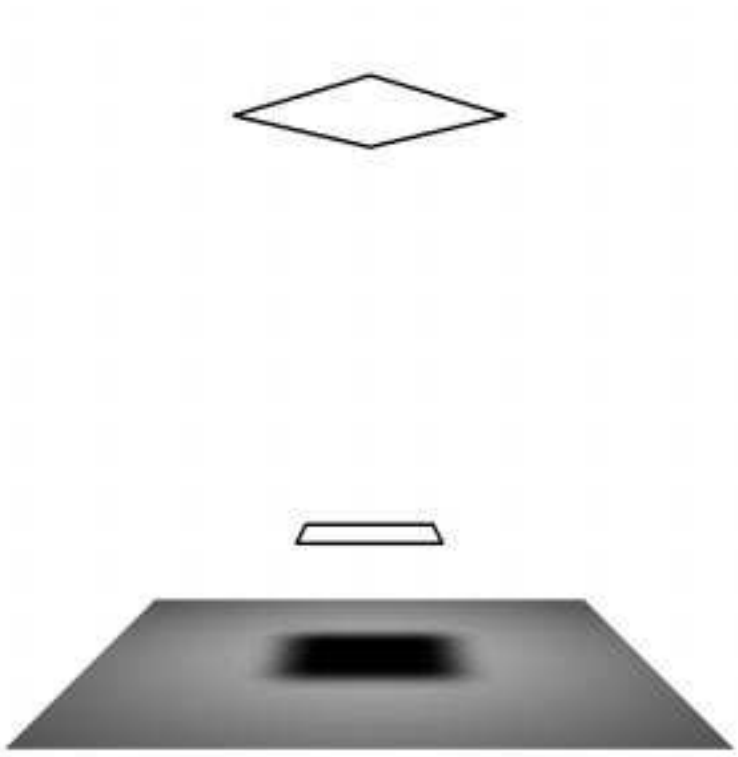
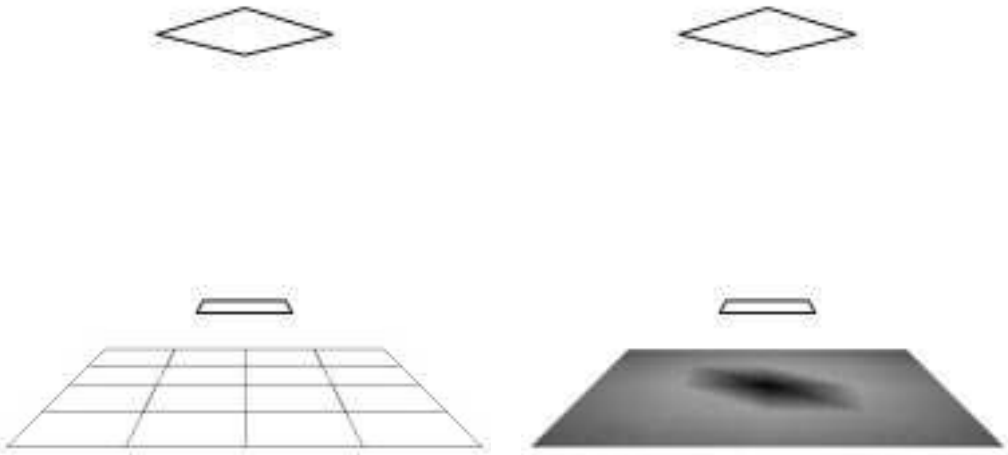


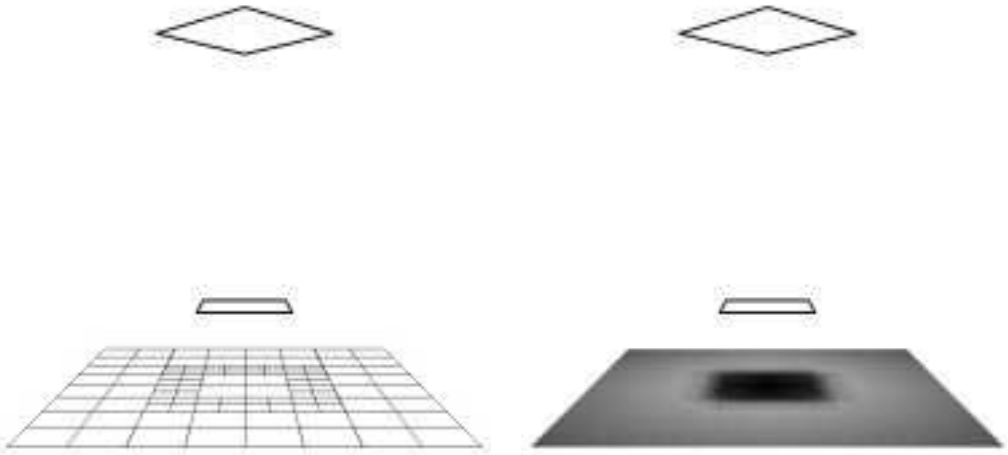
FIG. 7.13 – *La scène de tests pour la visibilité partielle,* $\varepsilon = 0,0001$.

On voit sur le tableau 7.4 le nombre de facettes produits par les trois algorithmes de raffinement en fonction des valeurs de ε, séparées en nombre de facettes en situation de visibilité totale, pour lesquelles on applique le critère établi à la section précédente, et facettes à visibilité partielle, pour lesquelles on emploie le critère établi ici. On voit ici que l'encadrement précis de la radiosité utilisé n'entraîne qu'une augmentation de l'ordre de 50 % du nombre de facettes en situation de visibilité partielle.

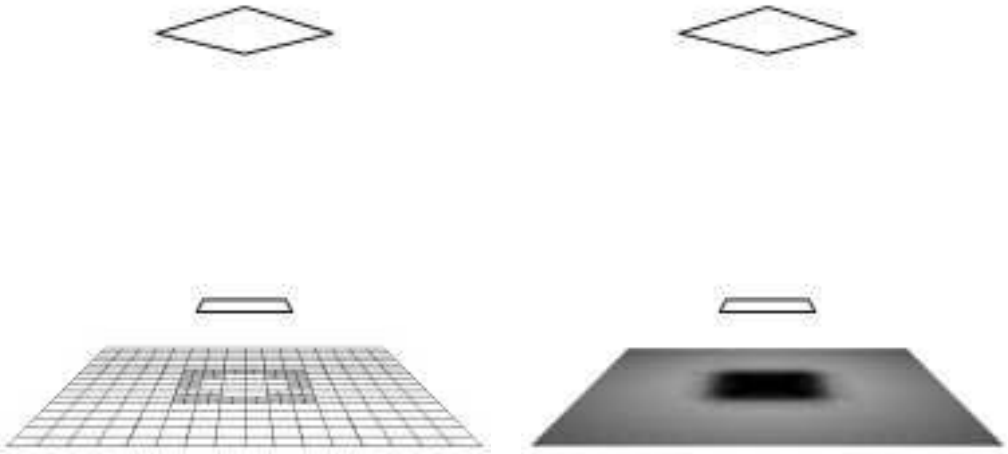
Le surcoût induit par notre algorithme en situation de visibilité partielle est lié aux



$\varepsilon = 1,0$

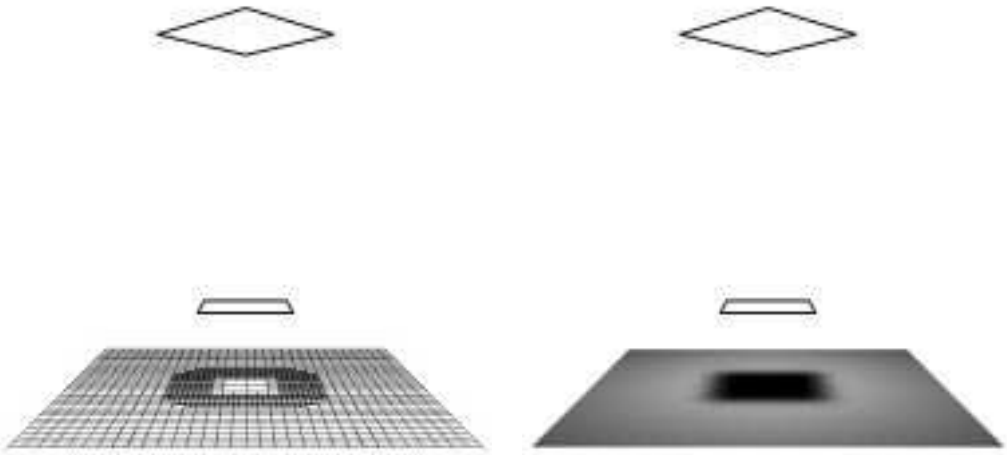


$\varepsilon = 0,1$

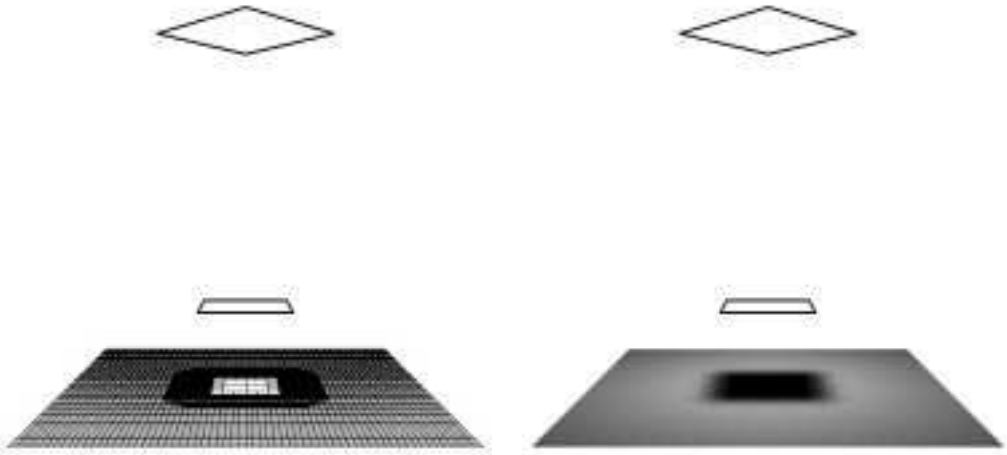


$\varepsilon = 0,01$

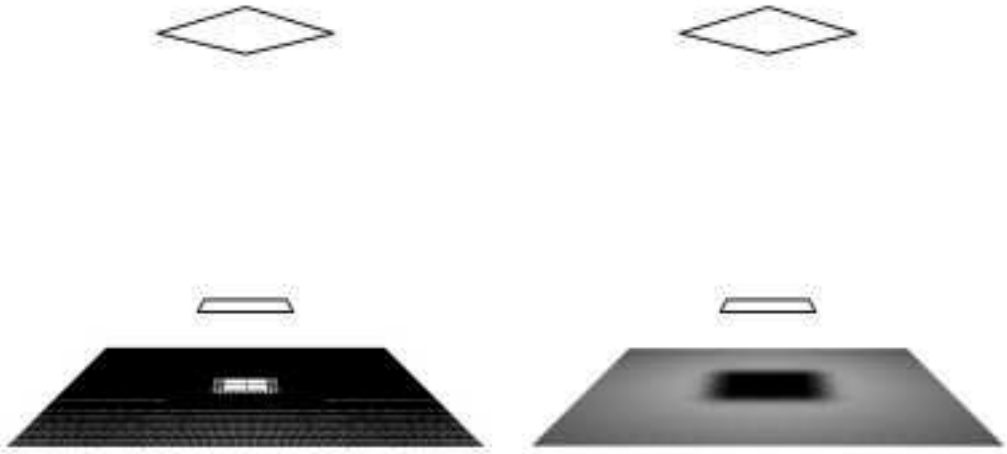
FIG. 7.14 – *Les images produites par les trois algorithmes lorsqu'ils coïncident.*



$\varepsilon = 0,001$

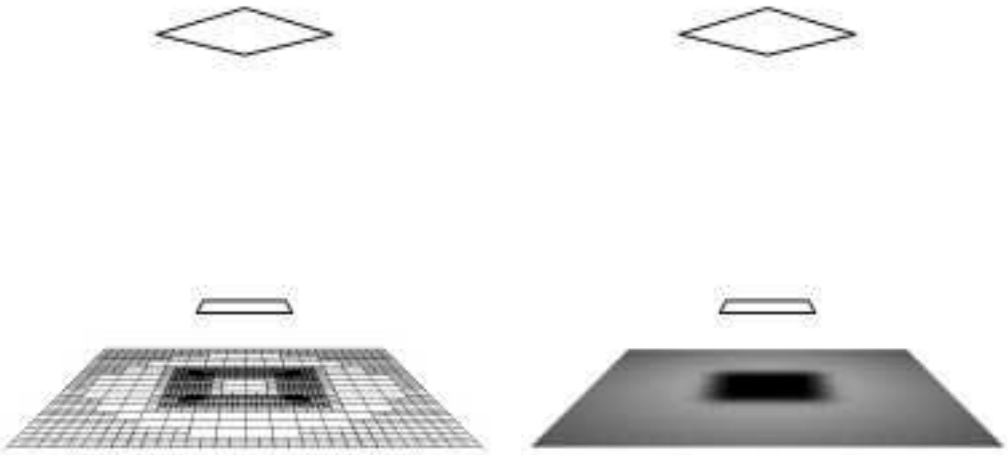


$\varepsilon = 0,0001$

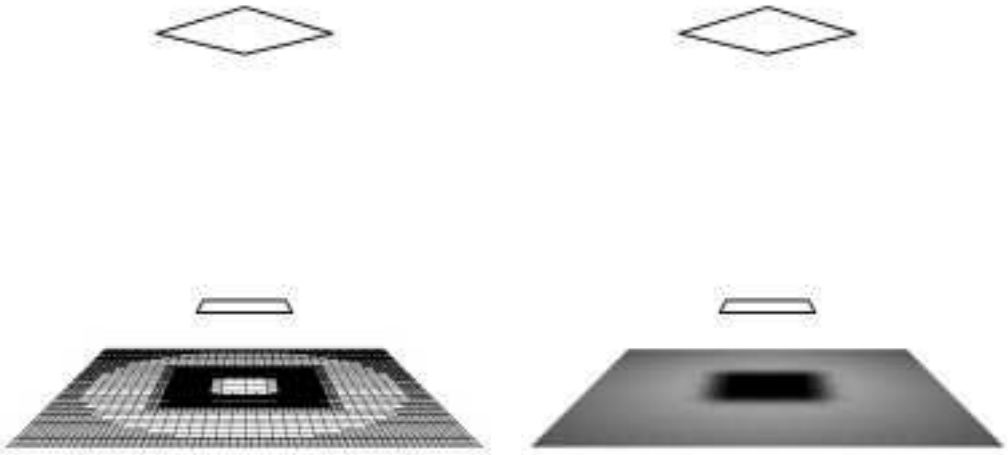


$\varepsilon = 0,00001$

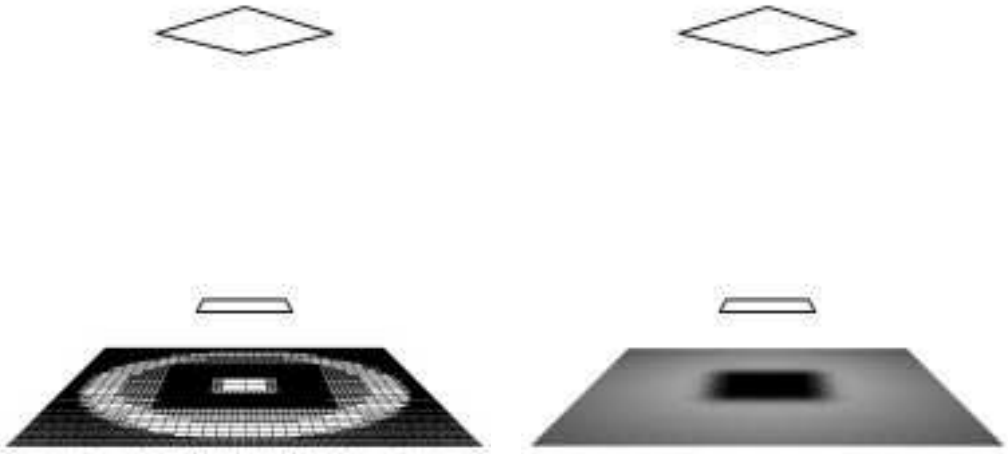
FIG. 7.15 – *Les images produites par l'algorithme heuristique.*



$\varepsilon = 0,001$



$\varepsilon = 0,0001$



$\varepsilon = 0,00001$

FIG. 7.16 – *Les images produites par l'algorithme de raffinement simplifié pour les mêmes valeurs de* ε.

TAB. 7.4 – *Nombre de facettes produits par les critères de raffinement en présence d'obstacles.*

	Visibilité totale				Visibilité partielle		
ε	Heuris.	Complet	Simplifié	Gain	Heuris.	Complet	Simplifié
1	12	12	12	0%	9	9	9
0,1	60	60	60	0 %	73	73	73
0,01	252	252	188	25 %	181	181	181
0,001	1032	1020	684	33 %	905	1221	1221
0,0001	4252	4216	2648	37 %	3445	6049	6049
0,00001	17180	17000	8888	48 %	13261	14977	14225

trois calculs d'occlusion de l'émetteur par l'obstacle que nous devons faire pour chaque facette au lieu d'un seul dans le cas de l'algorithme heuristique. Le temps de calcul est donc au maximum multiplié par trois. On constate que le critère de raffinement conduit à classer plus de facettes comme partiellement visibles que le critère heuristique. Le critère simplifié réduit le nombre de facettes employées dans la zone totalement visible de près de 50 %, et réduit même, d'un petit nombre, le nombre de facettes partiellement visibles pour ε suffisamment petit. Pour qu'il y ait réduction dans la zone partiellement visible, il faut que l'émetteur minimal et l'émetteur maximal soient très proches, en terme de distance géométrique, l'un de l'autre, ce qui ne peut se produire que dans les zones où une seule arête de l'obstacle cache l'émetteur. On voit d'ailleurs sur les maillages de la figure 7.16 que le raffinement est d'avantage poussé dans les zones de visibilité partielle correspondant à un coin de l'obstacle que dans les zones correspondant à une arête.

Nous avons également calculé la différence énergétique entre les différentes solutions liées aux maillages calculés et une solution de référence, le maillage complet avec $\varepsilon = 0,00001$. La différence locale d'énergie, calculée de la même manière que section 7.1.6 est donnée tableau 7.5. Nous n'avons pas fait figurer dans ce tableau le critère de raffinement avec simplification, qui donnait les mêmes valeurs que le critère complet.

En comparant les tableaux 7.5 et 7.2, on constate que l'erreur sur l'énergie a baissé en valeur, ce qui est normal puisque l'énergie totale de la scène a également baissé, du fait de la présence de l'obstacle. On observe par ailleurs que la précision de l'algorithme heuristique est moindre que celle de l'algorithme complet : pour $\varepsilon = 0,001$, par exemple, celui-ci, avec seulement 30 % de facettes supplémentaires réduit l'erreur locale d'un facteur deux.

Enfin, nous avons calculé pour chaque algorithme de raffinement, et pour chaque valeur de ε la valeur de l'énergie totale de la scène. Cette valeur de l'énergie est ensuite comparée avec l'énergie d'une scène de référence. Ce critère permet de mesurer l'erreur globale ; les résultats sont donnés tableau 7.6. Nous avons pris comme solution de référence le maillage produit par le critère complet pour $\varepsilon = 0,00001$.

Que peut-on tirer du tableau 7.6 ? Tout d'abord, que la mesure de l'énergie globale

TAB. 7.5 – *Différence locale de l'énergie en situation de visibilité partielle.*

	Heuristique		Complet	
ε	Facettes	Erreur	Facettes	Erreur
1,0	9	0,0110687	9	0,0110687
0,1	73	0,00199773	73	0,00199773
0,01	181	0,00199773	181	0,00199773
0,001	905	0,00011538	1221	$6{,}08327\times10^{-5}$
0,0001	3445	$8{,}04206\times10^{-6}$	6049	$6{,}94314\times10^{-6}$
0,00001	13261	$4{,}04035\times10^{-7}$	14977	0

TAB. 7.6 – *Mesure de l'énergie totale en présence d'obstacles.*

	Heuristique		Complet		Simplifié	
ε	Énergie	Diff.	Énergie	Diff.	Énergie	Diff.
1	7,536199	2,724 %	7,536199	2,724 %	7,536199	2,724 %
0,1	7,348277	0,162 %	7,348277	0,162 %	7,348277	0,162 %
0,01	7,306236	0,411 %	7,306236	0,411 %	7,295764	0,553 %
0,001	7,331510	0,066 %	7,331953	0,06 %	7,328719	0,1 %
0,0001	7,335487	0,012 %	7,335384	0,013 %	7,334480	0,026 %
0,00001	7,336278	0,001 %	7,336374	0 %	7,335583	0,01 %

n'est pas adaptée comme critère d'estimation de l'erreur dans une scène où la visibilité n'est pas totale ; en effet, une possible surestimation de la radiosité dans les zones où la visibilité est partielle peut compenser une sous-estimation dans les zones où la visibilité est totale, venant ainsi réduire la différence constatée au niveau de l'énergie. Ainsi, pour $\varepsilon = 0,1$, l'erreur calculée est inférieure à son niveau pour $\varepsilon = 0,01$.

De plus, une erreur importante en pourcentage commise dans une zone de visibilité partielle ou nulle a moins d'impact sur la valeur finale de l'énergie qu'une erreur plus faible commise dans une zone où la visibilité est totale. C'est ce qui explique que, bien que les maillages de discontinuité des critères complets et simplifiés soient égaux, l'erreur mesurée par notre algorithme pour le critère avec simplification du maillage est supérieure à l'erreur mesurée pour l'algorithme heuristique.

En résumé, nous avons présenté un algorithme capable de calculer un encadrement exact de la radiosité à partir de la construction d'émetteurs convexes encadrant l'émetteur original. Cet algorithme a été intégré au sein d'un algorithme de raffinement progressif, permettant ainsi de contrôler l'erreur commise sur chaque interaction.

Cet algorithme permet de réduire l'erreur commise dans la modélisation de la radiosité, et ce, sans augmenter de façon dramatique le nombre de facettes créées.

7.3 Intégration au sein d'un algorithme de radiosité hiérarchique

Les algorithmes que nous avons vu aux sections précédentes de calcul de l'erreur commise sur une interaction peuvent être inclus dans un algorithme de calcul de la radiosité de façon hiérarchique, étendant l'algorithme original de radiosité hiérarchique, dû à Hanrahan en 1990 [16], en définissant un nouveau critère de raffinement.

7.3.1 Modélisation de la radiosité sur les objets

Dans ce cadre, nous devons avoir une modélisation de la radiosité sur les objets qui puisse être employée à plusieurs niveaux. Comme nous utilisons des facteurs de forme ponctuels, les valeurs de la radiosité seront calculées aux sommets du maillage.

Lors de l'affichage des radiosités, nous utiliserons les valeurs de la radiosité stockées sur les sommets du maillage. De la sorte, il y a un contrôle des valeurs qui seront affichées au sein même du critère de raffinement.

Un exemple de méthode de radiosité utilisant des facteurs de forme ponctuels pour calculer les valeurs de la radiosité sur les sommets se trouve dans Max et Allison, 1992 [28].

Lors de l'émission de la radiosité, nous utilisons des liens entre objets, afin de pouvoir quantifier l'interaction et l'erreur commise dans sa modélisation. Ceci nous impose de connaître une valeur moyenne de la radiosité sur chaque facette, ainsi qu'une valeur maximale et une valeur minimale, afin de pouvoir conduire des encadrements.

On notera ici que nous n'utilisons pas la radiosité de façon symétrique lors de l'émission et lors de la réception : en effet, il est plus facile de traiter d'une radiosité constante lors de l'émission, alors qu'il est simple d'utiliser une radiosité variant de façon linéaire lors de la réception. Cette asymétrie n'a pas d'influence sur le contrôle de l'erreur ; elle le rend même possible plus facilement. Nous verrons plus loin (section 7.4.1) que l'on peut aussi envisager un contrôle de l'erreur dans le cadre d'une

radiosité linéaire à l'émission comme à la réception.

Notons enfin qu'une facette partage toutes les informations concernant ses sommets avec les facettes voisines. Il semble donc logique de partager ces informations dans la structure de données. Cependant, les problèmes qui se posent lorsqu'on partage les valeurs de radiosité entre facettes voisines – notamment lors de la phase de *push-pull* – font qu'il peut être préférable de stocker des valeurs de radiosité indépendantes pour chaque sommet suivant la facette à laquelle on considère qu'il appartient. Nous supposerons donc que pour chaque facette, on connait ses sommets et les valeurs de la radiosité aux sommets.

Outre les informations concernant ses sommets, une facette doit pouvoir être raffinée. Elle porte donc éventuellement la liste de ses facettes filles. Le nombre de facettes filles dépend de l'algorithme de maillage utilisé : il sera de 4 si l'on utilise un arbre quadtree, mais de 2 si l'on utilise un arbre binaire, tel un BSP-Tree. Il peut être éventuellement variable.

Lorsqu'on raffine une facette, il faut pouvoir vérifier si les facettes voisines ont déjà été raffinées, afin de récupérer les informations sur les sommets déjà calculés. Pour pouvoir répondre rapidement à ces requêtes, on stocke pour chaque facette la liste des facettes voisines. Cette liste des voisines est actualisée lorsque l'on raffine une facette, à la fois pour la facette raffinée et pour ses facettes voisines. Deux facettes sont voisines si elles partagent une arête.

Enfin, une facette doit porter la liste des objets avec lesquels elle interagit ; la nature de ces interactions et la structure de donnée employée pour les modéliser est l'objet de la section suivante. Lors de la phase de propagation de l'énergie le long de ces liens, seule la radiosité des facettes effectivement liées sera actualisée, comme dans l'algorithme de radiosité hiérarchique original, de Hanrahan [16]. Il faut donc prévoir une phase de propagation de l'énergie au sein de la hiérarchie, ou *push-pull*, dans laquelle la radiosité des facettes est mise à jour en fonction de la radiosité des facettes filles et des facettes parentes.

Dans l'algorithme de radiosité hiérarchique original, le principe du *push-pull* est simple : chaque facette mère envoie sa radiosité à ses facettes filles, et reçoit la moyenne de la radiosité de ses facettes filles. Cette mise à jour se fait par un parcours en profondeur d'abord de la hiérarchie de facettes. La structure de donnée de l'algorithme original utilise deux valeurs de la radiosité pour chaque facette : d'une part la radiosité à émettre, et d'autre part la radiosité qui a été reçue. Lors de la phase de *push-pull*, on actualise la radiosité à émettre en fonction de la radiosité qui a été reçue. Comme les facettes ne partagent aucune information, le *push-pull* peut se faire après un nombre arbitraire de propagations de l'énergie le long des liens.

Dans notre structure de données, une partie des informations est portée au niveau de la facette : la radiosité maximale et la radiosité minimale, par exemple. Pour ces valeurs, on peut employer une version légèrement modifiée de l'algorithme de *push-pull* : la radiosité minimale de la facette mère est ajoutée à la radiosité minimale calculée pour les facettes filles dans la phase de *push*, puis la radiosité minimale sur la facette mère est initialisée comme le minimum des radiosités sur les facettes filles lors de la phase de *pull*. On procède de la même manière pour la radiosité maximale.

La partie la plus délicate d'un algorithme de radiosité hiérarchique basé sur des valeurs de radiosité calculées aux sommets des facettes est l'adaptation de la procédure de *push-pull*. Cette adaptation dépend entièrement de la structure de données des sommets

et des facettes.

– Lorsque les valeurs de radiosité aux sommets ne sont pas partagées entre les facettes, l'algorithme de *push-pull* s'adapte de façon immédiate : lors de la phase de *push*, pour chaque facette, si elle est raffinée, on calcule les valeurs de radiosité pour les nouveaux sommets par une interpolation bilinéaire des valeurs des sommets de la facette. On ajoute aux valeurs de radiosité des sommets de la facette fille ces valeurs interpolées. Lors de la phase de *pull*, chaque facette reçoit de sa facette fille les nouvelles valeurs de la radiosité sur le ou les sommets qu'elles ont en commun.

– Si les valeurs des radiosités aux sommets sont partagées entre les facettes, on parvient à réduire l'encombrement mémoire, mais l'adaptation de la phase de *push-pull* est plus difficile. En effet, pour chaque sommet, il faudrait connaître la radiosité qu'il a reçu en tant que sommet faisant partie d'une facette, puisque cette radiosité sera utilisée pour actualiser la radiosité des sommets des facettes filles. Si on commence par un *push-pull* sur une facette, il faut prendre garde à conserver les valeur anciennes de la radiosité pour ne pas fausser le *push-pull* sur les facettes voisines.

Une fois qu'on a calculé les radiosités aux sommets, si l'on souhaite calculer la radiosité moyenne sur les facettes, on procède par une phase de *pull* supplémentaire : on calcule la radiosité sur les feuilles de la hiérarchie comme la moyenne des radiosités aux sommets de la facette feuille. Ensuite, en remontant vers la racine de l'arbre, la radiosité de chaque facette est prise égale à la moyenne des radiosités des facettes filles.

L'ensemble des structures de données utilisées pour modéliser la radiosité sont résumées figure 7.17.

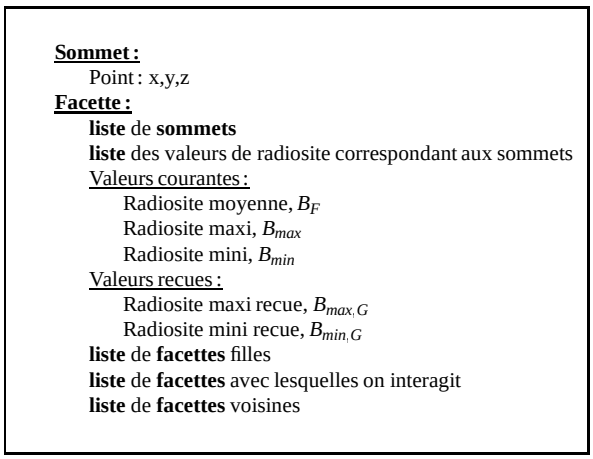


FIG. 7.17 – *Structures de données employées pour modéliser la radiosité.*

7.3.2 Modélisation et raffinement des interactions entre objets

L'interaction entre deux facettes est enregistrée sous la forme d'un lien. Ce lien est stocké sur la facette émettrice, au sein d'une liste de liens. Il porte l'identité de la facette

réceptrice, et un lien facette-sommet pour chacun des sommets de la facette réceptrice.

Lorsqu'un objet interagit avec un autre et qu'ils sont en situation de visibilité totale, la décision de raffiner ou non l'interaction se fera à partir des valeurs de la radiosité, du gradient et du Hessien dues à l'objet émetteur sur les sommets de l'objet récepteur. Si l'on décide de raffiner le récepteur, chaque facette fille créée aura sans doute l'un des sommets du récepteur parmi ses sommets. De plus, si les facettes voisines de la facette réceptrice interagissent également avec l'émetteur, elles partagent un des sommets de la facette réceptrice. Il serait souhaitable de pouvoir conserver l'information sur la radiosité, le gradient et le Hessien aux sommets, plutôt que de devoir la recalculer.

Cette information ne dépend que de la géométrie respective de la facette émettrice et du point de réception. La radiosité, le gradient et le Hessien sont proportionnels à la radiosité de l'émetteur, aussi peut-on ne stocker que la partie géométrique de cette information, le facteur de forme ponctuel pour la radiosité, et les équivalents pour le gradient et le Hessien.

Toutes ces informations purement géométriques sont regroupées dans une structure de données spécifique, le lien facette-sommet. Un lien facette-sommet peut être partagé par plusieurs interactions, c'est pourquoi il faut qu'il porte le nombre d'utilisations qui est fait de lui. À chaque fois qu'on crée un lien facette-facette utilisant ce lien facette-sommet, le nombre d'utilisations augment de un. À chaque fois qu'on détruit un tel lien, le nombre d'utilisations diminue de un. Lorsque le nombre d'utilisations vaut zéro, on peut détruire le lien facette-sommet.

Lorsque les deux objets qui interagissent ne sont pas en situation de visibilité totale, il faut stocker l'information concernant la visibilité : la nature des obstacles et suffisamment d'information pour pouvoir les retrouver. Cette information peut être une liste des obstacles efficaces, comme dans Teller, 1993 [42], ou un *cluster*, regroupant plusieurs objets, comme dans Sillion, 1995 [38]. Il est inutile de stocker l'émetteur maximal ou l'émetteur minimal, puisque cette information n'est pas réutilisable par d'autre facettes.

Enfin, chaque lien facette-facette porte aussi la valeur minimum et la valeur maximum du facteur de forme ponctuel entre la facette émettrice et la facette réceptrice.

La structure de donnée utilisée pour modéliser les interactions entre facettes est résumée figure 7.18.

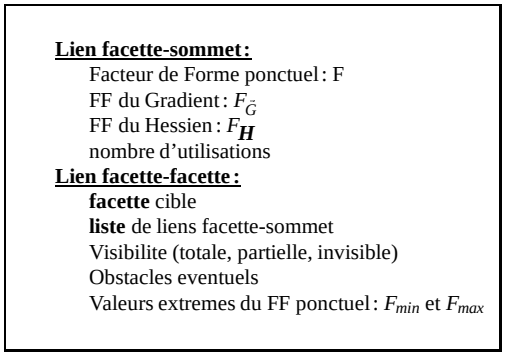


FIG. 7.18 – *Structures de données employées pour modéliser les interactions.*

Lors de la phase de raffinement des liens, on parcourt pour chaque facette les liens

qui en partent. Pour chaque lien facette-facette, en utilisant les valeurs stockées dans ses liens facette-sommet et la radiosité de l'émetteur, on détermine le maximum et le minimum de la radiosité qu'il induit sur la facette cible, en utilisant les techniques décrites section 7.1 pour les interactions avec visibilité totale et section 7.2 pour les interactions avec visibilité partielle. Si cet écart dépasse un certain seuil, comme à la section 7.1.6, on raffine le lien. Ce qui se fait en raffinant soit la facette cible, soit la facette émettrice. Dans les deux cas, on détruit l'ancien lien facette-facette, pour le remplacer par autant de liens facette-facette que la facette raffinée a de filles.

Si l'on raffine la facette cible, une partie des liens facette-sommets peut être réutilisée, permettant ainsi un gain de temps et de place mémoire : ainsi, si l'on raffine en utilisant un quadtree, la facette raffinée est remplacée par 4 facettes filles, ce qui correspond potentiellement à 9 liens facette-sommets (voir figure 7.19). De ces 9 liens, 4 ont déjà été calculés et n'ont pas besoin de l'être à nouveau. De même, si les facettes voisines de la facette cible ont déjà été raffinées – ce que l'on sait en utilisant la liste des liens vers les facettes voisines – une partie de ces liens a pu être calculée antérieurement. Dans le meilleur des cas, un seul lien facette-facette devra être recalculé.

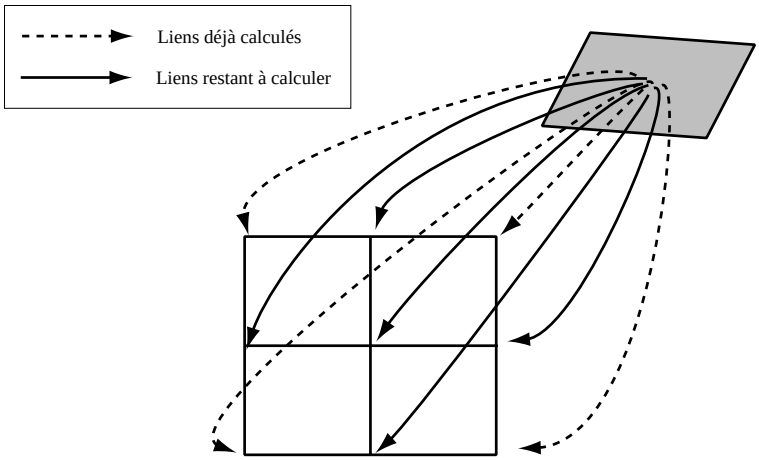


FIG. 7.19 – *Lors du raffinement d'un lien facette-facette, seule une partie des liens facette-sommets est à recalculer.*

Le code correspondant à la modélisation des interactions est résumé figure 7.20.

7.3.3 Propagation de l'énergie

Une fois établis les liens modélisant les interactions entre facettes à la précision souhaitée, on propage l'énergie suivant ces liens : pour chaque facette, on prend la liste des liens qui en partent, et on met à jour la radiosité des sommets des facettes cibles. En même temps qu'on met à jour la radiosité des sommets de la facette, on met à jour les valeurs minimum et maximum de la radiosité sur la facette, en utilisant les valeurs calculées pour le lien facette-facette.

Une fois que toutes les facettes ont propagé leur énergie, on effectue le *push-pull*, par un parcours en profondeur d'abord de la hiérarchie de facettes.

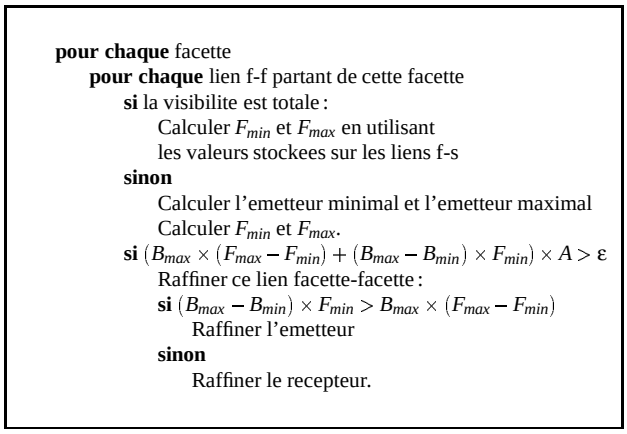


FIG. 7.20 – *Modélisation des interactions.*

Le code correspondant à la propagation de l'énergie dans la structure de données est résumé figure 7.21.

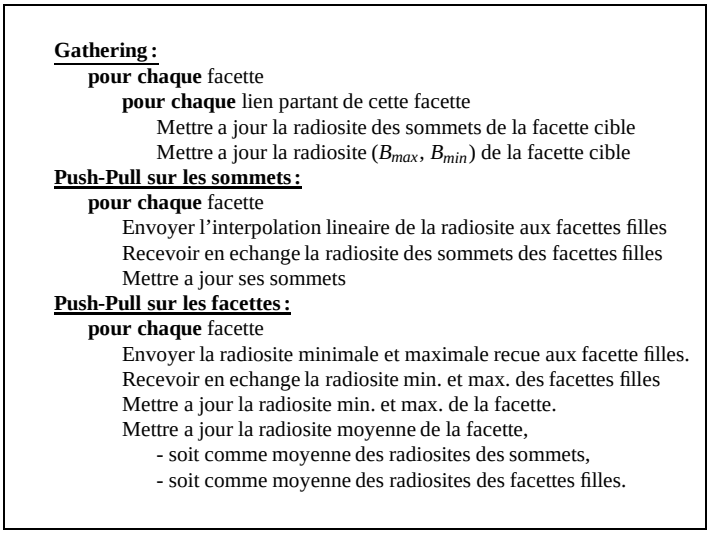


FIG. 7.21 – *Propagation de l'énergie.*

7.3.4 Contrôle de l'erreur de discrétisation

L'algorithme présenté ici contrôle l'erreur commise localement, sur chaque interaction. Afin de pouvoir contrôle également l'erreur de discrétisation, suivant les algorithmes décrits au chapitre 6, nous introduisons, après toutes les phases de propagation de l'énergie et de *push-pull*, une phase de contrôle de la cohérence interne. Dans cette phase, on compare la radiosité maximale et la radiosité minimale calculées sur chacune des facettes feuilles. Si l'encadrement est trop large, nous sommes en contradiction avec

notre hypothèse de radiosité constante sur une facette. Dans ce cas, on raffine la facette concernée, ainsi que tous les liens qui y aboutissent.

Cette situation peut se produire lorsqu'une source de lumière est décomposée en multiples facettes qui chacune individuellement sont trop petites pour que le facteur de forme justifie le raffinement des liens, mais qui prises ensemble ont un effet important. Un exemple de telle scène correspond à une ampoule lumineuse décomposée en multiples facettes.

Pour chaque facette, on calcule également la somme des facteurs de forme qui en partent, c'est-à-dire la somme des facteurs de forme portés par cette facette, et des facteurs de forme portés par les parents de cette facette et de la moyenne des facteurs de forme portés par les facettes filles. Si cette différence s'écarte de 1 au delà d'un certain seuil, et que nous sommes en présence d'une scène fermée, l'énergie quittant cette facette est mal modélisée. Pour y remédier, nous raffinons cette facette et tous les liens qui en partent.

L'ensemble des étapes de notre algorithme de radiosité hiérarchique modifié est rappelé figure 7.22.

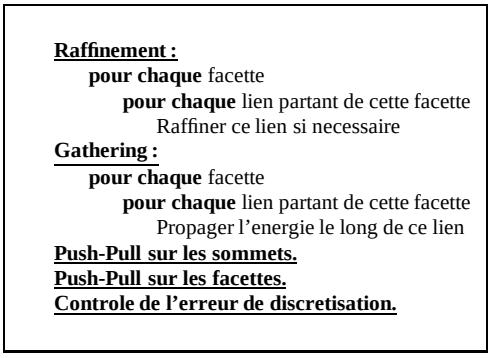


FIG. 7.22 – *Algorithme complet pour une itération.*

7.4 Améliorations éventuelles de l'algorithme

7.4.1 Radiosité linéaire sur l'émetteur

Notre algorithme de radiosité hiérarchique est asymétrique dans sa structure : la radiosité qui est utilisée lors des phases de propagation et de raffinement est une radiosité qui est supposée constante, alors que la radiosité qui sera affichée est une interpolation bilinéaire des valeurs aux sommets. Nous verrons, au chapitre 10, qu'il est possible de calculer la radiosité et ses dérivées lorsque la radiosité varie de façon linéaire sur l'émetteur, avec un encadrement aussi précis que l'on veut.

Dans ce cadre, il est envisageable de calculer la radiosité sur le récepteur en utilisant une approximation linéaire de la radiosité sur l'émetteur. On aurait alors d'une part une approximation linéaire de la radiosité sur le récepteur, propagée en utilisant les outils décrits au chapitre 10, et d'autre part une valeur constante correspondant à la différence entre cette approximation linéaire et la réalité, que l'on propagerait en utilisant les outils décrits au chapitre 8.

7.4.2 Utilisation du Hessien en présence d'obstacles

Notre algorithme de raffinement des interactions où la visibilité est partielle commence par déterminer un émetteur minimal et un émetteur maximal, convexes, pour pouvoir se ramener au cas de la visibilité totale en utilisant les conjectures de concavité du chapitre 5.

Nous avons vu au chapitre 5 que la radiosité en présence d'un obstacle dont le complémentaire est convexe semble se comporter comme la radiosité en l'absence d'obstacles. En particulier, dans ce cas, les conjectures de concavité semblent être vérifiées. Si l'on peut calculer le Hessien en présence d'obstacles – au moins en présence d'obstacles dont le complémentaire est convexe – on peut envisager une simplification du maillage produit par notre algorithme.

On calculerait pour chaque facette en situation de visibilité partielle le nombre d'arêtes de l'obstacle qui interviennent. Si une seule arête intervient, alors nous sommes dans un cas simple où le complémentaire est convexe. On n'a alors plus besoin de calculer l'émetteur minimal et l'émetteur maximal, mais uniquement les valeurs de la radiosité, du gradient et du Hessien aux sommets de la facette. Si la radiosité est concave sur la facette – c'est-à-dire si le déterminant du Hessien $rt - s^2$ est positif aux sommets de la facette – alors la radiosité de la facette est encadrée par un plan sécant et les plans tangents de la même manière qu'à la section 7.1.5.

7.5 Conclusion

Nous avons présenté un algorithme de radiosité hiérarchique implémentant un contrôle total de l'erreur commise sur chaque interaction, et un début de contrôle de l'erreur de discrétisation. Cet algorithme devrait permettre de modéliser avec précision chaque interaction, et de limiter tant le suréchantillonnage que le sous-échantillonnage inhérents à d'autres critères de raffinement.

Ce contrôle de l'erreur permet ainsi une meilleure adéquation des ressources à fournir en fonction de la précision souhaitée sur le résultat.

Troisième partie

Le gradient et le Hessien de la radiosité

8.

Cas des émetteurs constants : calcul du Jacobien et du Hessien

NOUS AVONS vu comment la connaissance des dérivées successives de la radiosité peut être mise à profit pour accélérer les calculs de radiosité. Il nous reste à voir comment l'on peut effectivement calculer ces dérivées successives, soit explicitement dans des cas simples, soit avec des approximations dans les cas plus complexes. Nous allons tout d'abord analyser le cas d'un émetteur polygonal sur lequel la radiosité est constante et qui est en situation de visibilité totale.

8.1 Travaux antérieurs

Le début de ce chapitre établit des résultats déjà publiés par Arvo, en 1994 [3], en les démontrant suivant une autre méthode. Nous allons voir que cette méthode permet de calculer également les dérivées successives, notamment le Hessien. Une partie de ce travail a déjà fait l'objet d'une publication, au *Sixth Eurographics Workshop on Rendering*, en 1995 [24].

8.2 L'expression de la radiosité

Nous connaissons une expression intégrale de la radiosité en un point x due à un émetteur A_2 (voir section 4.8 et figure 8.1) ; l'intégrale porte sur les contours de A_2.

$$B(x) = E(x) - B_0 \frac{\rho}{2\pi} \vec{n}_1 \cdot \oint_{\partial A_2} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 \tag{8.1}$$

Lorsque l'émetteur est un polygone (voir figure 8.1), l'intégrale sur les contours se transforme en une somme finie sur les arêtes du polygone :

$$B(x) = E(x) + \frac{\rho}{2\pi} B_0 \vec{n}_1 \cdot \sum_i \vec{\gamma}_i \tag{8.2}$$

$\vec{\gamma}_i$ est le vecteur de norme l'angle sous lequel l'arête i du polygone est vue du point x, et de direction le produit vectoriel $\vec{r}_i \times \vec{r}_{i+1}$. Voir section 8.4, page 137 la démonstration de ceci.

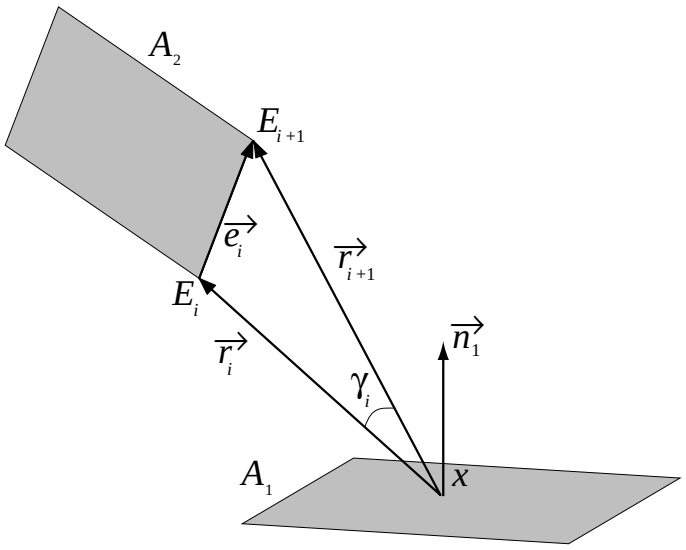


FIG. 8.1 – *Cas où l'émetteur est un polygone.*

8.3 Le gradient de la radiosité

On a ainsi trois formules distinctes (équations 4.6, 4.7 et 8.2) pour exprimer la radiosité en un point x sous l'influence d'un émetteur polygonal.

Chacune de ces trois formules est une expression tout à fait acceptable de la radiosité. On peut dériver formellement chacune des formules pour avoir une expression du gradient de radiosité.

Pour calculer le gradient d'une fonction, on peut soit revenir à la définition du gradient, et calculer la dérivée de f dans une direction $\vec{v}$, pour ensuite chercher si les dérivées sont liées par une application linéaire. On peut aussi utiliser les relations définissant le gradient des fonctions composées et des produits de fonctions.

Lorsqu'elle est applicable, la deuxième méthode est plus simple et plus rapide.

De même, il est plus simple et plus rapide pour le calcul du gradient d'utiliser la définition de la fonction de radiosité basée sur une intégrale de contour.

Formellement, le gradient de radiosité s'écrit alors :

$$\nabla B(x) = \nabla E(x) - B_0 \frac{\rho}{2\pi} \nabla \left(\vec{n}_1 \cdot \oint_{\partial A_2} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 \right)$$

Habituellement, les variations dues à $E(x)$ sont connues. Ce sont davantage les variations de la radiosité reçue qui nous intéressent. De plus, dans la majeure partie des cas, E est constante. Dans toute la suite, nous prendrons $\nabla E(x) = 0$.

Lorsque la surface sur laquelle on intègre (A_2) ne dépend pas de la position du point où l'on calcule la radiosité, c'est-à-dire lorsqu'il n'y a pas d'obstacles entre le récepteur et l'émetteur, on peut tirer le gradient sous l'intégrale.

On a alors :

$$\nabla B(x) = -B_0 \frac{\rho}{2\pi} \oint_{\partial A_2} \nabla \left[\vec{n}_1 \cdot \left(\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 \right) \right] \tag{8.3}$$

Pour aller plus loin, il nous faut étudier le gradient d'un produit mixte dont un seul terme est variable ; soit donc $\nabla \left[\vec{a} \cdot \left(\vec{b} \times \vec{c} \right) \right]$, avec $\vec{a}$ et $\vec{b}$ constants. Nous utiliserons les règles qui décrivent le gradient d'un produit scalaire, et le rotationnel d'un produit vectoriel.

$$\begin{aligned} \nabla \left[\vec{a} \cdot \left(\vec{b} \times \vec{c} \right) \right] &= \vec{a} \times \left[\nabla \times \left(\vec{b} \times \vec{c} \right) \right] + (\vec{a} \cdot \nabla) \left(\vec{b} \times \vec{c} \right) \\ &= \vec{a} \times \left[(\vec{c} \cdot \nabla) \vec{b} \right] - \left(\nabla \cdot \vec{b} \right) (\vec{a} \times \vec{c}) + \left[(\vec{a} \cdot \nabla) \vec{b} \right] \times \vec{c} \end{aligned} \tag{8.4}$$

À ce point, il nous faut calculer $\nabla \cdot \vec{b}$ et $(\vec{a} \cdot \nabla)\vec{b}$, avec $\vec{a}$ quelconque, et $\vec{b} = \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2}$. Notons que, contrairement aux habitudes, nous dérivons par rapport à l'origine du vecteur $\vec{r}_{12}$. D'où un signe moins.

$$\nabla \cdot \left(\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \right) = -\frac{1}{\|\vec{r}_{12}\|^2} \tag{8.5}$$

$$\vec{a} \cdot \nabla \left(\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \right) = 2\frac{\vec{r}_{12}(\vec{a} \cdot \vec{r}_{12})}{\|\vec{r}_{12}\|^4} - \frac{\vec{a}}{\|\vec{r}_{12}\|^2} \tag{8.6}$$

En reportant les résultats des équations 8.5 et 8.6 dans l'équation 8.4, il vient :

$$\begin{aligned} \nabla \left[\vec{a} \cdot \left(\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times \vec{c} \right) \right] &= \vec{a} \times \left(2\frac{\vec{r}_{12}(\vec{c} \cdot \vec{r}_{12})}{\|\vec{r}_{12}\|^4} - \frac{\vec{c}}{\|\vec{r}_{12}\|^2} \right) + \frac{\vec{a} \times \vec{c}}{\|\vec{r}_{12}\|^2} \\ &+ \left(2\frac{\vec{r}_{12}(\vec{a} \cdot \vec{r}_{12})}{\|\vec{r}_{12}\|^4} - \frac{\vec{a}}{\|\vec{r}_{12}\|^2} \right) \times \vec{c} \\ &= -\frac{\vec{a} \times \vec{c}}{\|\vec{r}_{12}\|^2} + 2\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \left((\vec{a} \cdot \vec{r}_{12})\vec{c} - (\vec{c} \cdot \vec{r}_{12})\vec{a} \right) \\ &= -\frac{\vec{a} \times \vec{c}}{\|\vec{r}_{12}\|^2} + 2\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \left((\vec{a} \times \vec{c}) \times \vec{r} \right) \\ &= -\frac{\vec{a} \times \vec{c}}{\|\vec{r}_{12}\|^2} + 2\frac{\vec{r}_{12} \cdot \vec{r}_{12}(\vec{a} \times \vec{c})}{\|\vec{r}_{12}\|^4} - 2\frac{\vec{r}_{12} \cdot (\vec{a} \times \vec{c})\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \\ &= \frac{\vec{a} \times \vec{c}}{\|\vec{r}_{12}\|^2} + 2\frac{\vec{a} \cdot (\vec{r}_{12} \times \vec{c})\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \end{aligned}$$

En reportant ce dernier résultat dans l'équation 8.3, nous avons l'expression du gradient de radiosité :

$$\nabla B(x) = -B_0 \frac{\rho}{2\pi} \left(\vec{n}_1 \times \oint_{\partial A_2} \frac{d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} + 2 \oint_{\partial A_2} \frac{\vec{n}_1 \cdot (\vec{r}_{12} \times d\vec{\ell}_2)\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \right) \tag{8.7}$$

8.4 Calculs et implémentation

Nous avons (équations 4.7 et 8.7) une définition de la radiosité comme une intégrale de contour, et une définition du gradient comme une intégrale de contour. Dans le cas où l'émetteur est un polygone, ces deux définitions se simplifient.

Notons E_i le i^{e} sommet du polygone émetteur, $\vec{r}_i$ le vecteur liant x à E_i, $\vec{e}_i$ le vecteur reliant E_i à E_{i+1} (voir figure 8.1). Nous avons alors une définition du gradient et de la radiosité comme des sommes finies :

$$\begin{aligned} B(x) &= E(x) - B_0 \frac{\rho}{2\pi} \vec{n}_1 \cdot \sum_i \int_{E_i}^{E_{i+1}} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 \\ \nabla B(x) &= -B_0 \frac{\rho}{2\pi} \sum_i \vec{n}_1 \times \int_{E_i}^{E_{i+1}} \frac{d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} + 2 \int_{E_i}^{E_{i+1}} \frac{\vec{n}_1 \cdot (\vec{r}_{12} \times d\vec{\ell}_2)\vec{r}_{12}}{\|\vec{r}_{12}\|^4} \end{aligned}$$

Notons que, si l'on paramètre chaque arête par $y \in [0,1]$, on a :

$$\begin{aligned} \vec{r}_{12} &= \vec{r}_i + y\vec{e}_i \\ d\vec{\ell}_2 &= \vec{e}_i dy \\ \vec{r}_{12} \times d\vec{\ell}_2 &= \vec{r}_i \times \vec{e}_i dy \\ &= \vec{r}_i \times \vec{r}_{i+1} dy \end{aligned}$$

Ce paramétrage permet d'exprimer B et ∇B en fonction d'intégrales simples :

$$\begin{aligned} B(x) &= E(x) - B_0 \frac{\rho}{2\pi} \sum_i \vec{n}_1 \cdot (\vec{r}_i \times \vec{r}_{i+1}) \int_0^1 \frac{dy}{\|\vec{r}_{12}\|^2} \\ \nabla B(x) &= -B_0 \frac{\rho}{2\pi} \sum_i \vec{n}_1 \times \vec{e}_i \int_0^1 \frac{dy}{\|\vec{r}_{12}\|^2} \\ &+ 2\vec{n}_1 \cdot (\vec{r}_i \times \vec{r}_{i+1}) \int_0^1 \frac{\vec{r}_i + y\vec{e}_i}{\|\vec{r}_{12}\|^4} dy \end{aligned}$$

On choisit de noter :

$$\begin{aligned} I_n(x,i) &= \int_0^1 \frac{dy}{\|\vec{r}_{12}\|^{2n}} \\ J_n(x,i) &= \int_0^1 \frac{y dy}{\|\vec{r}_{12}\|^{2n}} \end{aligned}$$

Un bref calcul donne les valeurs de $I_n(x,i)$ et $J_n(x,i)$, pour $n \neq 1$, d'abord :

$$\begin{aligned} I_n(x,i) &= \frac{1}{2(n-1)} \frac{1}{\|\vec{e}_i \times \vec{r}_i\|^2} \left(\frac{\vec{e}_i \cdot \vec{r}_{i+1}}{r_{i+1}^{2(n-1)}} - \frac{\vec{e}_i \cdot \vec{r}_i}{r_i^{2(n-1)}} + (2n-3) e_i^2 I_{n-1} \right) \\ J_n(x,i) &= \frac{1}{2(n-1)e_i^2} \left(\frac{1}{r_i^{2(n-1)}} - \frac{1}{r_{i+1}^{2(n-1)}} \right) - \frac{\vec{r}_i \cdot \vec{e}_i}{e_i^2} I_n \end{aligned}$$

Et pour $n = 1$:

$$\begin{aligned} I_1(x,i) &= \frac{\gamma_i}{\|\vec{e}_i \times \vec{r}_i\|} \\ J_1(x,i) &= \frac{1}{2e_i^2} \ln\left(\frac{e_{i+1}}{e_i}\right) - \frac{\vec{e}_i \cdot \vec{r}_i}{e_i^2} I_1 \end{aligned}$$

Notons que ces valeurs sont purement géométriques, et peuvent se calculer en fonction des vecteurs $\vec{e}_i$ et des vecteurs $\vec{r}_i$.

On a donc, en fonction de I_n et J_n,

$$\begin{aligned} B(x) &= E(x) - B_0 \frac{\rho}{2\pi} \sum_i \vec{n}_1 \cdot (\vec{r}_i \times \vec{r}_{i+1})\, I_1 \\ \nabla B(x) &= -B_0 \frac{\rho}{2\pi} \sum_i \vec{n}_1 \times \vec{e}_i I_1 + 2\vec{n}_1 \cdot (\vec{r}_i \times \vec{r}_{i+1})\, (\vec{r}_i I_2 + \vec{e}_i J_2) \end{aligned}$$

L'implémentation se fait très facilement en `C++`, en utilisant une classe vecteur implémentant les opérateurs usuels, tels que produit scalaire et produit vectoriel, et en recopiant mot pour mot la formule mathématique.

8.5 Étude de complexité

La plupart des termes utilisées pour calculer la radiosité peuvent être réutilisés pour le calcul du gradient. Le plus efficace consiste à calculer le gradient et la radiosité en même temps, pour réutiliser le plus grand nombre possible de termes.

Un point important avant d'envisager d'utiliser le gradient de radiosité est de chiffrer le surcoût qu'il entraîne par rapport au calcul de la radiosité seule. Pour cela, il faut dénombrer toutes les opérations requises par les calculs de l'un et de l'autre. Ces opérations sont de natures différentes : additions, soustractions, racines carrées, fonctions trigonométriques inverses... Nous allons donc devoir comparer des racines carrées et des additions. C'est bien sûr impossible, à moins de savoir à combien d'additions équivaut un calcul de racine carrée.

On peut faire effectuer dans une boucle un grand nombre d'additions à un ordinateur donné, comparer avec la même boucle, mais vide et en déduire le temps moyen d'une addition. On peut faire de même pour les autres opérateurs, et exprimer pour chaque opérateur utilisé (racines carrées, arc cosinus, multiplications...) le nombre d'additions auquel il équivaut. Ce qui nous donne des grandeurs que l'on peut alors comparer.

Ces travaux sont forcément reliés à un type d'ordinateur et à un type de compilateur. Ils sont également très peu précis. Néanmoins, ils permettent de comparer les temps de calcul relatifs du gradient et de la radiosité. Ils ont également l'avantage d'être basés sur l'expérience, et non sur des valeurs théoriques.

TAB. 8.1 – *Temps de calcul comparés des opérateurs utilisés.*

Ordinateur	$+$	$-$	$\times$	$\div$	$\sqrt{}$	arccos
SGI, par défaut	1	1	1	1	23	40
SGI, `-mips2`	1	1	1	1	8	40
LC 630	1	1	1	1	4,5	40

Les valeurs équivalentes trouvées se trouvent dans le tableau 8.1. J'ai utilisé d'une part des stations de travail Silicon Graphics avec le compilateur standard, et d'autre

part un Macintosh LC630, sans coprocesseur arithmétique. Remarquons que les temps relatifs sont les mêmes pour toutes les stations Silicon Graphics à base de processeur R4000 (Indy, Indigo 2). Naturellement, les temps absolus sont sensiblement différents. Pour les SGI, les temps de calcul comparés sont différents si l'on fait appel à une optimisation utilisant le code spécifique du R4000 (option `-mips2`) et si l'on utilise les options par défaut, qui génèrent du code compatible avec les anciens processeurs.

Le résultat le plus important est que les quatre opérations arithmétiques prennent le même temps sur tous les ordinateurs testés. Nous ne les distinguerons donc plus dans la suite. Un produit scalaire est ainsi équivalent à 5 opérations, un produit vectoriel à 9 opérations.

8.5.1 La complexité de la radiosité

On peut écrire γ_i comme $\arccos\left(\frac{\vec{e}_i \cdot \vec{r}_i}{e_i r_i}\right)$. Le calcul de γ_i nécessite donc deux normes, un produit scalaire, une division, et un arc cosinus ; ce qui fait 16 opérations, deux racines carrées et un arc cosinus.

Le calcul de $I_1(x,i)$ nécessite un produit vectoriel, une norme et une division en sus de γ_i. Soit au total, 31 opérations, 3 racines carrées et 1 arc cosinus.

Le calcul de $B(x)$ nécessite n calculs de $I_1(x,i)$ – en supposant un émetteur polygonal à n côtés – n produits mixtes $\vec{n}_1 \cdot (\vec{r}_i \times \vec{r}_{i+1})$, et $2n+3$ opérations.

Donc, pour calculer $B(x)$:

- $47n+3$ opérations,
- $3n$ racines carrées,
- n arc cosinus.

8.5.2 La complexité du gradient de radiosité

Pour calculer $\nabla B(x)$, il faut $I_1(x,i)$, $I_2(x,i)$ et $J_2(x,i)$. $I_1(x,i)$ a déjà été calculé lors du calcul de $B(x)$.

Tous les produits scalaires et produits vectoriels nécessaires pour calculer $I_2(x,i)$ ont déjà été calculés pour la radiosité, sauf $\vec{e}_i \cdot \vec{r}_{i+1}$. I_2 ne coûte donc que 12 opérations supplémentaires.

De même, $J_2(x,i)$ réutilise uniquement des produits scalaires déjà calculés, et donc ne coûte que 6 opérations.

Calculer le gradient de radiosité conjointement avec la radiosité exige n calculs de $I_2(x,i)$ et $J_2(x,i)$, et $28n+2$ opérations en sus pour effectuer la sommation.

Le coût supplémentaire induit par le calcul du gradient est ainsi de $46n+2$ opérations.

Ainsi, calculer le gradient en même temps que la radiosité double sensiblement le nombre d'opérations à effectuer, mais ne nécessite aucun calcul de racines carrées ou de fonctions trigonométriques. Réaliser les calculs conjointement ne fera, dans le pire des cas, que doubler le temps de calcul.

Mais le fait que la radiosité nécessite plusieurs calculs de racines carrées et de fonctions trigonométriques inverses fait encore diminuer le coût relatif du gradient. Le tableau 8.2 montre, pour les ordinateurs étudiés dans le tableau 8.1, le nombre équivalent

TAB. 8.2 – *Coûts comparés du gradient et de la radiosité (nombre d'additions équivalent).*

Ordinateur	$B(x)$	$B(x)$ et $\nabla B(x)$	Surcoût
SGI, par défaut	$156n+3$	$202n+5$	30 %
SGI, `-mips2`	$111n+3$	$157n+5$	42 %
LC 630	$100n+3$	$146n+5$	46 %

d'additions nécessaires pour calculer la radiosité et le gradient, ainsi que le surcoût engendré par le calcul du gradient. Les surcoûts ont été calculés en prenant pour hypothèse des émetteurs triangulaires ($n = 3$).

On voir ainsi que, pour les ordinateurs testés, le calcul du gradient conjointement avec la radiosité ne représente qu'un surcoût de l'ordre de 40%. C'est-à-dire que l'expression directe que nous avons présenté ici, par dérivation de la formule intégrale de la radiosité et intégration de la formule du gradient obtenu est beaucoup plus rapide qu'un calcul par différences finies qui chercherait plusieurs valeurs de radiosité en des points proches de x pour en déduire le gradient.

8.6 Implémentation et résultats

Pour tester la validité de notre implémentation, nous avons calculé la radiosité et le gradient en tous les points d'un récepteur plan donné, sous l'influence d'un émetteur carré (voir figures 8.2 et 8.3). Le gradient, qui est un vecteur, ne peut pas être visualisé. En revanche, on peut visualiser sa norme ; c'est ce qui est fait figure 8.4.

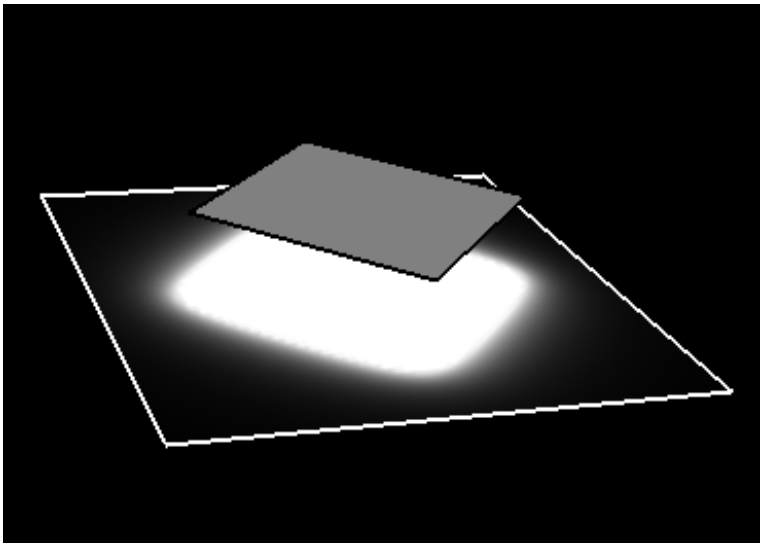


FIG. 8.2 – *La géométrie de notre scène de tests.*

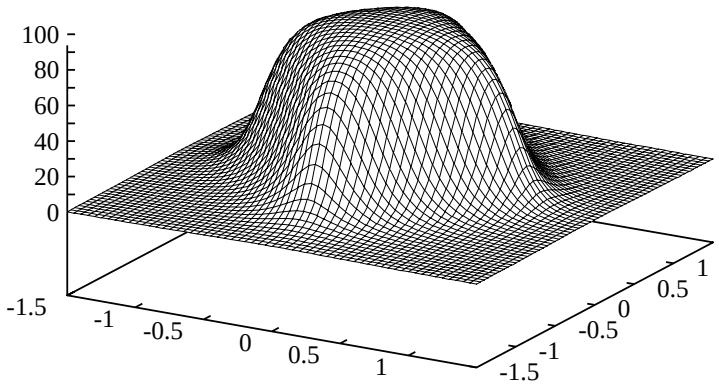


FIG. 8.3 – *La radiosité sur le récepteur.*

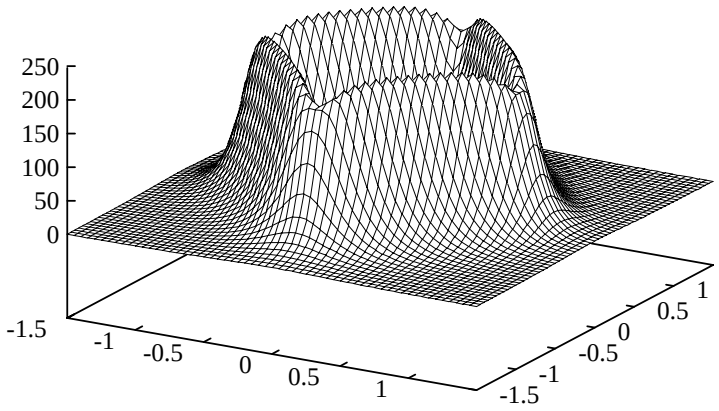


FIG. 8.4 – *La norme du gradient de la radiosité sur le récepteur.*

Pour tester la qualité du gradient calculé, nous avons également intégré le champ vectoriel du gradient. Le résultat est montré figure 8.5. La différence entre la radiosité calculée et le résultat de l'intégration du gradient est montrée dans les figure 8.6.

Dans notre scène de test, l'émetteur est placé très près du récepteur, pour produire des variations rapides de la radiosité et du gradient. Ces variations permettent de pousser l'algorithme d'intégration et celui de calcul du gradient vers leurs limites. De plus, l'émetteur étant parallèle au récepteur, on a une connaissance intuitive de ce que devrait être la forme de la radiosité, ce qui fournit un autre moyen de vérifier la validité des calculs.

8.7 Le Hessien de la radiosité

Comme nous l'avons vu au chapitre 4.7, le calcul pratique du Hessien se fait en dérivant formellement deux fois l'expression de la radiosité, suivant un vecteur $\vec{v}$, puis

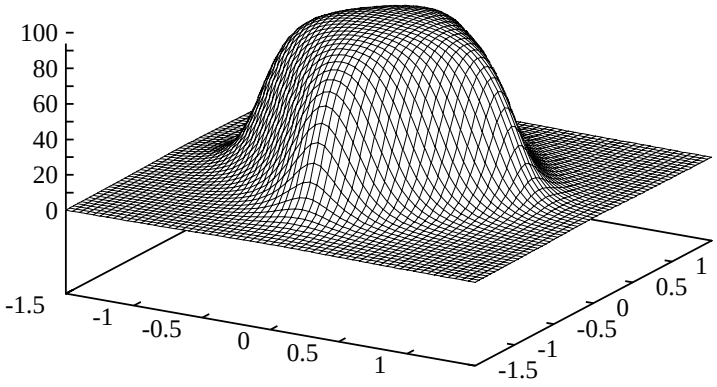


FIG. 8.5 – *Intégration du gradient sur le récepteur.*

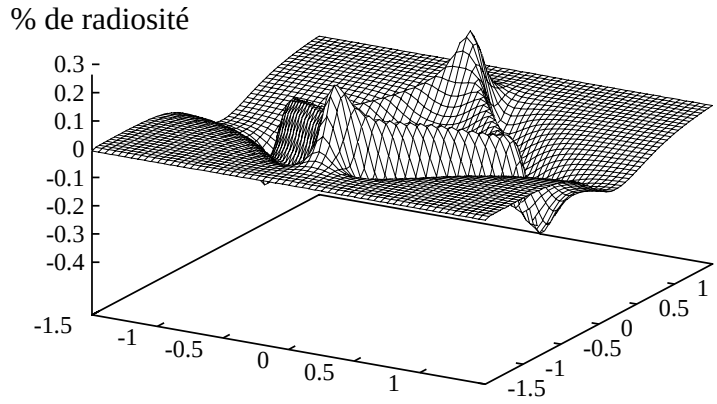


FIG. 8.6 – *Différence entre l'intégration du gradient et la radiosité.*

suivant un vecteur $\vec{u}$.

Si, au lieu d'utiliser l'expression de la radiosité comme une intégrale surfacique (équation 4.6) nous utilisons l'expression de la radiosité comme une intégrale de contour (équation4.7), nous avons :

$$\frac{\partial^2 B(x)}{\partial\vec{u}\partial\vec{v}} = \frac{\partial\nabla B(x)}{\partial\vec{u}}\cdot\vec{v}$$

$$\begin{aligned}\frac{\partial\nabla B(x)}{\partial\vec{u}} &= -\frac{2B_0\rho}{2\pi}\left(\oint_{\partial A_2}\frac{(\vec{n}_1\times d\vec{\ell}_2)(\vec{r}_{12}\cdot\vec{u})-\vec{n}_1\cdot(\vec{u}\times d\vec{\ell}_2)\vec{r}_{12}}{\|\vec{r}_{12}\|^4}\right.\\ &- \left.\oint_{\partial A_2}\frac{\vec{n}_1\cdot(\vec{r}_{12}\times d\vec{\ell}_2)}{\|\vec{r}_{12}\|^4}(\vec{u}\cdot\vec{v})+4\oint_{\partial A_2}\frac{\vec{n}_1\cdot(\vec{r}_{12}\times d\vec{\ell}_2)}{\|\vec{r}_{12}\|^6}(\vec{r}_{12}\cdot\vec{u})\vec{r}_{12}\right)\end{aligned}$$

En reprenant la notation $\boldsymbol{Q}(\vec{a},\vec{b})$ introduite en 4.7, on a l'expression du Hessien de la radiosité :

$$\begin{aligned}\boldsymbol{H} &= -\frac{2B_0\rho}{2\pi}\left(\oint_{\partial A_2}\boldsymbol{Q}\left(\vec{n}_1\times d\vec{\ell}_2,\frac{\vec{r}_12}{\|\vec{r}_{12}\|^4}\right)\right.\\ &+ \left.\oint_{\partial A_2}\frac{\vec{n}_1\cdot(\vec{r}_{12}\times d\vec{\ell}_2)}{\|\vec{r}_{12}\|^4}\left(2\frac{\boldsymbol{Q}(\vec{r}_{12},\vec{r}_{12})}{\|\vec{r}_{12}\|^2}-\boldsymbol{I}\right)\right)\end{aligned}$$

Si nous supposons un émetteur polygonal, le Hessien s'exprime comme une somme de termes (on reprend les notations de 8.4) :

$$\boldsymbol{H} = -\frac{2B_0\rho}{2\pi}\sum_i\int_0^1\boldsymbol{P}_i dy = -\frac{2B_0\rho}{2\pi}\sum_i\boldsymbol{H}_i$$

avec :

$$\begin{aligned}\boldsymbol{P}_i &= \boldsymbol{Q}\left(\vec{n}_1\times\vec{e}_i,\frac{\vec{r}_i+y\vec{e}_i}{\|\vec{r}_{12}\|^4}\right)+\frac{\vec{n}_1\cdot(\vec{r}_i\times\vec{e}_i)}{\|\vec{r}_{12}\|^4}\left(2\frac{\boldsymbol{Q}(\vec{r}_{12},\vec{r}_{12})}{\|\vec{r}_{12}\|^2}-\boldsymbol{I}\right)\\ \boldsymbol{Q}(\vec{r}_{12},\vec{r}_{12}) &= \boldsymbol{Q}(\vec{r}_i,\vec{r}_i)+y^2\boldsymbol{Q}(\vec{e}_i,\vec{e}_i)+2y\boldsymbol{Q}(\vec{r}_i,\vec{e}_i)\\ \boldsymbol{H}_i &= \boldsymbol{Q}\left(\vec{n}_1\times\vec{e}_i,\int_0^1\frac{\vec{r}_i+y\vec{e}_i}{\|\vec{r}_{12}\|^4}dy\right)-\boldsymbol{I}\vec{n}_1\cdot(\vec{r}_i\times\vec{e}_i)\int_0^1\frac{dy}{\|\vec{r}_{12}\|^4}\\ &+ 2\vec{n}_1\cdot(\vec{r}_i\times\vec{e}_i)\int_0^1\frac{1}{\|\vec{r}_{12}\|^6}\left(\boldsymbol{Q}(\vec{r}_i,\vec{r}_i)+y^2\boldsymbol{Q}(\vec{e}_i,\vec{e}_i)+2y\boldsymbol{Q}(\vec{r}_i,\vec{e}_i)\right)dy\\ \boldsymbol{H}_i &= \boldsymbol{Q}(\vec{n}_1\times\vec{e}_i,\vec{r}_iI_2+\vec{e}_iJ_2)-\vec{n}_1\cdot(\vec{r}_i\times\vec{e}_i)I_2\boldsymbol{I}\\ &+ 2\vec{n}_1\cdot(\vec{r}_i\times\vec{e}_i)\left(\boldsymbol{Q}(\vec{r}_i,\vec{r}_i)I_3+\boldsymbol{Q}(\vec{e}_i,\vec{e}_i)K_3+2J_3\boldsymbol{Q}(\vec{r}_i,\vec{e}_i)\right)\end{aligned}$$

On a repris la notation I_n, J_n introduite plus haut (section 8.4), et introduit par surcroît K_n :

$$\begin{aligned}K_n &= \int_0^1\frac{y^2dy}{\|\vec{r}_{12}\|^{2n}}\\ &= \frac{1}{e_i^2}\left(I_{n-1}-r_i^2I_n-2(\vec{r}_i\cdot\vec{e}_i)J_n\right)\end{aligned}$$

On voit que la plupart des termes de notre expression du Hessien s'expriment en fonction d'éléments déjà calculés lors du calcul de la radiosité. Une fois de plus, le calcul simultané de la radiosité, du gradient et du Hessien se fera plus rapidement que les calculs séparés.

8.8 Implémentation du Hessien

L'implémentation du Hessien se fait en recopiant la formule donnée plus haut. Les calculs sont plus coûteux que pour le gradient, puisqu'on doit effectuer des calculs matriciels : le calcul de $Q(\vec{r}_i, \vec{e}_i)$ représente ainsi 24 opérations, et celui de $Q(\vec{r}_i, \vec{r}_i)$ en représente 18. Chaque opération sur les matrices (addition, multiplication par un scalaire) représente 9 opérations, et ainsi de suite.

Le calcul du Hessien représente ainsi $138n + 11$ opérations pour un émetteur à n côtés. On diminue le nombre de calculs en effectuant d'abord les opérations sur des scalaires avant de travailler sur des matrices, par exemple en calculant d'abord $\vec{a} = 2\vec{n}_1 \cdot (\vec{r}_i \times \vec{e}_i) I_3 \vec{r}_i$, puis $\boldsymbol{Q}(\vec{a}, \vec{r}_i)$, ce qui coûte 24 opérations, plutôt que de calculer $\boldsymbol{Q}(\vec{r}_i, \vec{r}_i)$, puis de multiplier par I_3, puis par $2\vec{n}_1 \cdot (\vec{r}_i \times \vec{e}_i)$, ce qui coûterait 45 opérations.

Le tableau 8.3 donne les temps de calcul comparés de la radiosité seule et de la radiosité conjointement avec le gradient et le Hessien.

Approximativement, le temps de calcul du Hessien correspond au temps de calcul de la radiosité ; pour calculer ensemble la radiosité et ses deux premières dérivées, il faut donc $2,5$ fois plus de temps que pour calculer la radiosité seule.

TAB. 8.3 – *Coûts comparés du Hessien et de la radiosité (nombre d'additions équivalent).*

Ordinateur	$B(x)$	$B(x)$ et $\boldsymbol{H}_B(x)$	Surcoût
SGI, par défaut	$156n+3$	$340n+16$	120 %
SGI, `-mips2`	$111n+3$	$295n+16$	164 %
LC 630	$100n+3$	$284n+16$	186 %

Qui plus est, pour tirer du Hessien la valeur de la dérivée seconde dans une direction donnée $\vec{u}$, il faut calculer $\boldsymbol{H}\vec{u}$ (15 opérations), puis $\vec{u} \cdot (\boldsymbol{H}\vec{u})$ (5 opérations), soit un total de 20 opérations.

Enfin, si l'on veut savoir si la fonction de radiosité est concave en un point, il faut calculer deux vecteurs $\vec{u}$ et $\vec{v}$ orthonormés et contenus dans le plan, faire opérer le Hessien sur ces deux vecteurs pour en tirer r, s et t, puis calculer $rt - s^2$.

En reprenant la scène de tests de la figure 8.2, on peut calculer le Hessien en tous les points du récepteur. Ce Hessien ne peut pas être visualisé en tant que tel. On peut cependant visualiser son déterminant $rt - s^2$, qui est positif seulement sur un convexe, négatif partout ailleurs (voir figure 8.7). À l'aide de la ligne de niveau 0, on remarque que le déterminant est bien positif seulement sur un convexe. On remarque aussi de larges «pics» qui correspondent aux coins du carré, des points où courbure de la radiosité est très forte. En effet, le déterminant du Hessien est lié à la courbure de la surface.

On peut aussi utiliser le Hessien pour calculer la valeur de la dérivée seconde par rapport à x de la radiosité (voir figure 8.9). Le gradient nous donne la dérivée première (voir figure 8.8). On voit sur la figure 8.9 que sur une droite dirigée par x la dérivée seconde de la radiosité est positive, sauf sur un convexe borné où elle est négative.

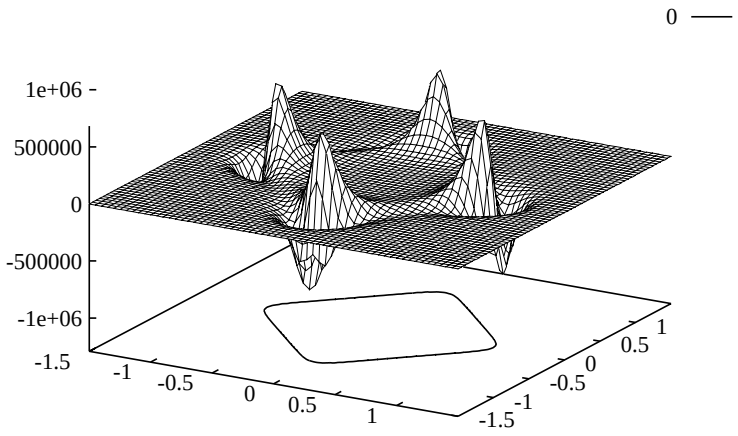


FIG. 8.7 – *Le déterminant du Hessien de la radiosité.*

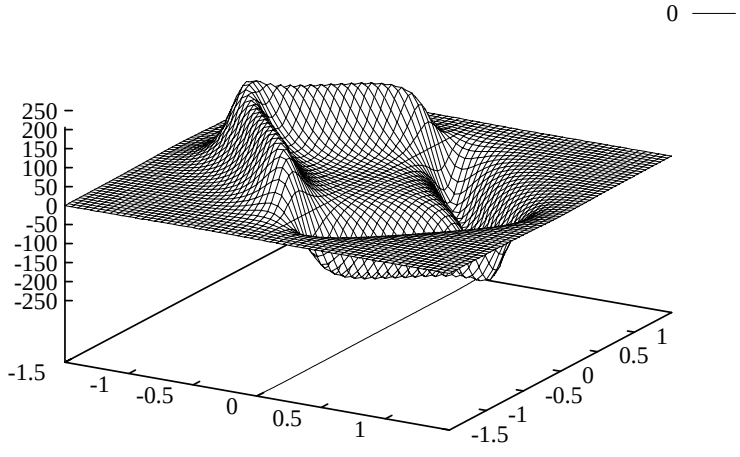


FIG. 8.8 – *La dérivée de la radiosité suivant x.*

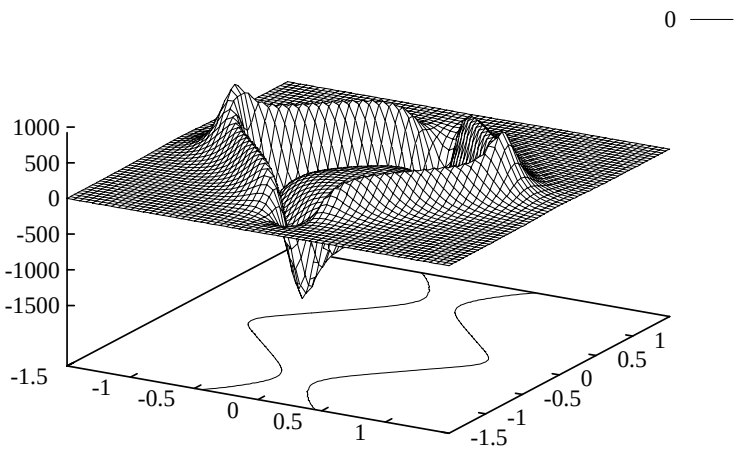


FIG. 8.9 – *La dérivée seconde de la radiosité suivant x.*

9.

Le gradient de radiosité en présence d'obstacles

LE CALCUL du gradient est – somme toute – fort simple en l'absence d'obstacles entre l'émetteur et le récepteur. Ici, nous nous intéressons au cas où des obstacles sont présents. Dès lors, la surface sur laquelle porte l'intégration (ou son contour) varie en fonction de la position du point récepteur (voir figure 9.1).

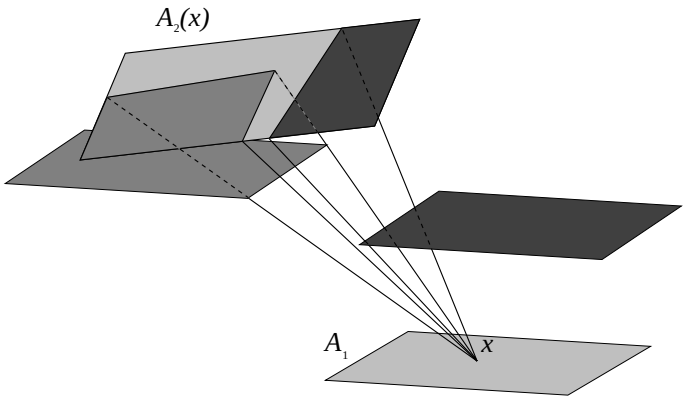


FIG. 9.1 – *Exemple d'émetteur partiellement bloqué.*

Nous allons voir maintenant comment les résultats du chapitre précédent s'étendent aux cas où il y a des obstacles. Ces résultats ont déjà été établis par Arvo, en 1994 [3] ; nous retrouvons ici les mêmes résultats, en utilisant une autre méthode. La section 9.5 étend cette méthode au calcul du Hessien de la radiosité en présence d'obstacles, un résultat entièrement nouveau.

9.1 Première étude, théorique

Ainsi qu'on l'a vu au chapitre 4, section 4.6, lorsque l'émetteur dépend du point x, le gradient de la radiosité s'écrit (voir équation 4.4) :

$$\nabla B(x) = \nabla E(x) + \frac{\rho}{\pi}\left[\int_{A_2(x)} B(y)\nabla\left(\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}\right)dA_2 + \boldsymbol{J}\left(\int_{A_2(x)}\right)B(y)\frac{\cos\theta_1\cos\theta_2}{\|r_{12}\|^2}dA_2\right]$$

Il est clair que le gradient de B, dans le cas général, s'écrit comme une somme de deux termes, le premier tenant compte de la variation de la fonction à intégrer, $I\nabla f$, et le deuxième tenant compte de la variation du domaine d'intégration, à cause des conditions de visibilité entre la source et le récepteur.

Nous avons étudié le premier terme au chapitre précédent, le deuxième terme reste à étudier. Cette étude théorique ne nous donne pas de formules explicites le concernant. Pour le connaître mieux, il faut se pencher sur l'expression du Jacobien de l'opérateur d'intégration :

$$\boldsymbol{J}\left(\int_{A_2(x)}\right) f$$

9.2 Étude de la dérivée de l'opérateur d'intégration

Il nous faut revenir à la définition de la dérivée suivant un vecteur ; considérons donc la fonction e :

$$e(h) = B(x+h\vec{v}) - B(x)$$

La valeur qui nous intéresse est celle de la dérivée de $B(x)$ suivant le vecteur $\vec{v}$, soit :

$$\lim_{h\to 0}\frac{1}{h}e(h)$$

Si nous notons $A_2(x)$ la surface de l'émetteur qui est visible du point x – celle sur laquelle porte l'intégration – nous avons :

$$e(h) = \int_{A_2(x+h\vec{v})} f(x+h\vec{v},y)\,dA_2 - \int_{A_2(x)} f(x,y)\,dA_2$$

La figure 9.1 montre un exemple de calcul de radiosité en présence d'obstacles, avec $A_2(x)$ et de $A_2(x+h\vec{v})$.

Ou encore :

$$\begin{aligned} e(h) &= \int_{A_2(x+h\vec{v})} f(x+h\vec{v},y)\,dA_2 \\ &- \int_{A_2(x+h\vec{v})} f(x,y)\,dA_2 \\ &+ \int_{A_2(x+h\vec{v})} f(x,y)\,dA_2 \\ &- \int_{A_2(x)} f(x,y)\,dA_2 \end{aligned} \tag{9.1}$$

Donc :

$$\begin{aligned} e(h) &= \int_{A_2(x+h\vec{v})-A_2(x)} f(x,y)dA_2 \\ &+ \int_{A_2(x+h\vec{v})} (f(x+h\vec{v},y)-f(x,y))\,dA_2 \end{aligned}$$

Si on introduit le gradient de f,

$$e(h) = \int_{A_2(x+h\vec{v})-A_2(x)} f(x,y)dA_2 + h\vec{v}\cdot\int_{A_2(x)} \nabla f(x,y)dA_2 + o(h)$$

On retrouve une somme de deux termes. Le deuxième terme est ce que nous avons étudié au chapitre précédent. Le premier est la variation de la radiosité due aux changements des conditions de visibilité, pour lequel il nous reste à trouver une formule explicite.

La partie inconnue, celle qui nous reste à calculer, est :

$$e_1(h) = \int_{A_2(x+h\vec{v})-A_2(x)} f(x,y)dA_2$$

En appliquant le théorème de Stokes (voir équation 4.1, page 60), nous avons :

$$\begin{aligned} e_1(h) &= \vec{n}_1 \cdot \int_{\partial A_2(x+h\vec{v})-\partial A_2(x)} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 \\ &= \vec{n}_1 \cdot \vec{e}_2(h) \end{aligned}$$

Le contour de $A_2(x+h\vec{v})$ peut se calculer à partir du contour de $A_2(x)$ et du Jacobien de chacun des sommets. Si nous notons E_i les sommets de $A_2(x)$, $\boldsymbol{J}_i$ leur Jacobien, les sommets de $A_2(x+h\vec{v})$ sont :

$$E'i = E_i + h\boldsymbol{J}_i\vec{v} + \vec{o}(h)$$

Nous avons donc une définition de $\vec{e}_1(h)$:

$$\vec{e}_1(h) = \vec{n}_1 \cdot \sum_i \int_{E'_i}^{E'_{i+1}} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 - \int_{E_i}^{E_{i+1}} \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2} \times d\vec{\ell}_2 + o(h) \qquad (9.2)$$

Sur le segment $[E_iE_{i+1}]$,

$$\begin{aligned} \vec{r}_{12} &= \overrightarrow{xE_i} + y\overrightarrow{E_iE_{i+1}} = \vec{r}_i + y\vec{e}_i \\ d\vec{\ell}_2 &= \vec{e}_i dy \\ \vec{r}_{12} \times d\vec{\ell}_2 &= \vec{r}_i \times \vec{e}_i dy \end{aligned}$$

Sur le segment $[E'_iE'_{i+1}]$,

$$\begin{aligned} \delta\boldsymbol{J}_i &= \boldsymbol{J}_{i+1} - \boldsymbol{J}_i \\ \vec{r}_{12} &= \overrightarrow{xE'_i} + y\overrightarrow{E'_iE'_{i+1}} \\ &= \vec{r}_i + y\vec{e}_i + h\boldsymbol{J}_i\vec{v}h(\boldsymbol{J}_{i+1}-\boldsymbol{J}_i)\vec{v} \\ &= \vec{r}_i + y\vec{e}_i + h(\boldsymbol{J}_i + y\delta\boldsymbol{J}_i)\vec{v} \\ d\vec{\ell}_2 &= (\vec{e}_i + h(\boldsymbol{J}_{i+1}-\boldsymbol{J}_i)\vec{v})dy \\ &= (\vec{e}_i + h\delta\boldsymbol{J}_i\vec{v})dy \\ \vec{r}_{12} \times d\vec{\ell}_2 &= \vec{r}_i \times \vec{e}_i + h(\vec{r}_i \times (\boldsymbol{J}_{i+1}-\boldsymbol{J}_i)\vec{v} - \vec{e}_i \times \boldsymbol{J}_i\vec{v}) \\ &= \vec{r}_i \times \vec{e}_i + h(\vec{r}_i \times \boldsymbol{J}_{i+1}\vec{v} - r_{i+1} \times \boldsymbol{J}_i\vec{v}) \end{aligned}$$

Un bref calcul montre que :

$$\begin{aligned}\lim_{h\to 0}\frac{1}{h}e_1(h) &= \sum_i I_1\left[(\vec{r}_{i+1}\times\vec{n}_1)\cdot\boldsymbol{J}_i\vec{v}-(\vec{r}_i\times\vec{n}_1)\cdot\boldsymbol{J}_{i+1}\vec{v}\right]\\ &+ 2\vec{n}_1\cdot(\vec{e}_i\times\vec{r}_i)\left[(\vec{r}_iI_2+\vec{e}_iK_2)\cdot\delta\boldsymbol{J}_i\vec{v}+(\vec{e}_i\cdot\boldsymbol{J}_i\vec{v}+\vec{r}_i\cdot\delta\boldsymbol{J}_i\vec{v})\,J_2\right]\end{aligned}$$

On a réintroduit la notation I_n, J_n utilisée à la section 8.4, et K_n introduite à la section 8.7.

Ce que nous voulons, c'est exprimer $\lim_{h\to 0}\frac{1}{h}e_1(h)$ sous la forme $\vec{G}\cdot\vec{v}$. De la sorte, $\vec{G}$ représentera la contribution des obstacles au gradient de $B(x)$. Le Jacobien des sommets $\boldsymbol{J}_i$ est une matrice 3×3. On peut remarquer que ce Jacobien n'apparaît que dans des expressions de la forme $\vec{a}\cdot(\boldsymbol{J}_i\vec{v})$. Par définition du produit scalaire et de la transposée,

$$\vec{a}\cdot(\boldsymbol{J}_i\vec{v}) = ({}^{\mathrm{t}}\boldsymbol{J}_i\vec{a})\cdot\vec{v}$$

Donc,

$$\begin{aligned}\vec{G} &= \sum_i I_1\left[{}^{\mathrm{t}}\boldsymbol{J}_i(\vec{r}_{i+1}\times\vec{n}_1)-{}^{\mathrm{t}}\boldsymbol{J}_{i+1}(\vec{r}_i\times\vec{n}_1)\right]\\ &+ 2\vec{n}_1\cdot(\vec{e}_i\times\vec{r}_i)\left[{}^{\mathrm{t}}\delta\boldsymbol{J}_i(\vec{r}_iI_2+\vec{e}_iK_2)+\left({}^{\mathrm{t}}\boldsymbol{J}_i\vec{e}_i+{}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{r}_i\right)J_2\right]\end{aligned}$$

Nous avons ainsi la contribution des obstacles au gradient de radiosité, en fonction des Jacobiens des sommets du polygone, ou plutôt de leur transposée. Voyons maintenant l'expression du Jacobien d'un sommet et de sa transposée.

9.3 Le Jacobien d'un sommet, étude des quatre types de sommets

Le Jacobien d'un sommet de $A_2(x)$ dépend de son origine. Distinguons dans le contour de $A_2(x)$ les arêtes réelles, qui sont les arêtes originales du polygone (même si elles ne sont pas parcourues de bout en bout) et les arêtes virtuelles qui sont les arêtes dues à la présence d'un obstacle entre le point de réception, x et l'émetteur (voir figure 9.2).

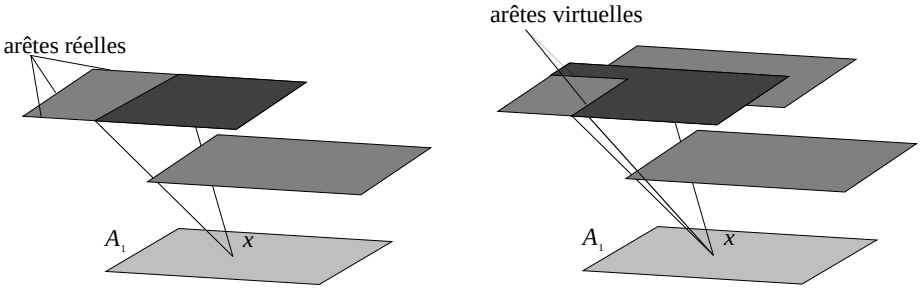


FIG. 9.2 – *Arêtes du contour de $A_2(x)$.*

On a alors quatre type de sommets de $A_2(x)$ (voir figure 9.3) :

1. l'intersection de deux arêtes réelles ;

2. l'intersection d'une arête réelle et d'une arête virtuelle ;
3. l'intersection de deux arêtes virtuelles provenant du même obstacle ;
4. l'intersection de deux arêtes virtuelles provenant de deux obstacles différents.

FIG. 9.3 – *Différents types de sommets.*

Le Jacobien d'un sommet est l'opérateur linéaire tel que :

$$\frac{dM}{d\vec{v}} = J\vec{v}$$

Pour connaître le Jacobien, il suffit donc de calculer la dérivée suivant un vecteur quelconque.

9.3.1 Type 1

Le premier cas, intersection de deux arêtes réelles, est réglé tout de suite : le sommet ne dépend pas de la position de x. Son Jacobien est donc nul.

9.3.2 Type 2

Le deuxième cas de figure est l'intersection d'une arête réelle et d'une arête virtuelle.

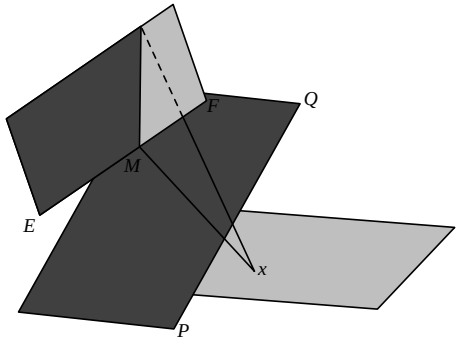


FIG. 9.4 – *Type 2 : intersection d'une arête réelle et d'une arête virtuelle.*

Une arête virtuelle est la projection sur l'émetteur de l'arête d'un obstacle (voir figure 9.4). Notons $[EF]$ l'arête réelle du polygone émetteur et $[PQ]$ l'arête de l'obstacle qui est à l'origine de l'arête virtuelle.

Le sommet, M, s'exprime comme :

$$M = E + \frac{\overrightarrow{xP} \cdot (\overrightarrow{PQ} \times \overrightarrow{EP})}{\overrightarrow{xP} \cdot (\overrightarrow{PQ} \times \overrightarrow{EF})} \overrightarrow{EF}$$

Et sa dérivée suivant la direction $\vec{v}$ est :

$$\frac{dM}{d\vec{v}} = \frac{\vec{v} \cdot (\overrightarrow{PQ} \times \overrightarrow{PF})}{[\overrightarrow{xO} \cdot (\overrightarrow{PQ} \times \overrightarrow{EF})]^2} \overrightarrow{EF}$$

Notons que cette dérivée peut s'exprimer sous la forme :

$$\frac{dM}{d\vec{v}} = (\vec{m} \cdot \vec{v}) \overrightarrow{EF} \tag{9.3}$$

9.3.3 Type 3

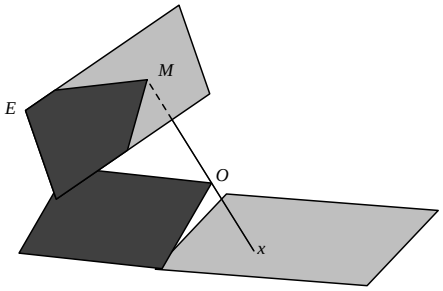


FIG. 9.5 – *Type 3 : intersection de deux arêtes virtuelles dues au même obstacle.*

Le troisième cas de figure est l'intersection de deux arêtes virtuelles provenant du même obstacle.

Ce sommet est la projection sur l'émetteur du sommet d'un obstacle. Notons M le sommet de type 3, P le sommet de l'obstacle qui en est l'origine. Notons $\vec{n}_2$ la normale à l'émetteur et d la distance à l'origine de l'émetteur (voir figure 9.5). Une équation de l'émetteur est alors : $\overrightarrow{OM} \cdot \vec{n}_2 = d$

Le sommet, M, s'exprime comme :

$$M = x + \frac{d - \overrightarrow{Ox} \cdot \vec{n}_2}{\overrightarrow{xP} \cdot \vec{n}_2} \overrightarrow{xP}$$

Et sa dérivée suivant la direction $\vec{v}$ est :

$$\frac{dM}{d\vec{v}} = \frac{\overrightarrow{OP} \cdot \vec{n}_2 - d}{(\overrightarrow{xP} \cdot \vec{n}_2)^2} \vec{n}_2 \times (\vec{v} \times \overrightarrow{xP})$$

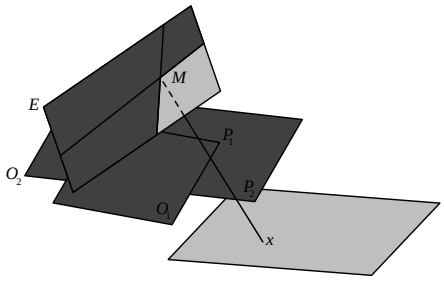


FIG. 9.6 – *Type 4 : intersection de deux arêtes virtuelles provenant de deux obstacles différents.*

9.3.4 Type 4

Le quatrième cas possible est l'intersection de deux arêtes virtuelles provenant de deux obstacles différents.

Chaque arête virtuelle est la projection sur l'émetteur de l'arête d'un obstacle (voir figure 9.6). Notons M le sommet obtenu, $[P_1Q_1]$ l'arête du premier obstacle, $[P_2Q_2]$ celle du deuxième obstacle. Notons comme précédemment $\vec{n}_2$ la normale à l'émetteur et d la distance à l'origine de l'émetteur.

On pose :

$$\begin{aligned}
\vec{e}_1 &= \overrightarrow{P_1Q_1} \\
\vec{e}_2 &= \overrightarrow{P_2Q_2} \\
\vec{m}_1 &= \overrightarrow{xP_1} \times \vec{e}_1 \\
\vec{m}_2 &= \overrightarrow{xP_2} \times \vec{e}_2 \\
\vec{u} &= \vec{m}_1 \times \vec{m}_2
\end{aligned}$$

Le sommet, M, s'exprime comme :

$$M = x + \frac{d - \overrightarrow{Ox} \cdot \vec{n}_2}{\vec{u} \cdot \vec{n}_2} \vec{u}$$

Et sa dérivée suivant la direction $\vec{v}$ est :

$$\begin{aligned}
\frac{dM}{d\vec{v}} &= \frac{\vec{n}_2 \times}{(\vec{u} \cdot \vec{n}_2)^2} \Big[(\vec{m}_2 \cdot \vec{v}) \vec{m}_1 (\vec{u} \cdot \vec{n}_2 - (\vec{m}_1 \cdot \vec{e}_2)(d - \overrightarrow{Ox} \cdot \vec{n}_2)) \\
&- (\vec{m}_1 \cdot \vec{v}) \vec{m}_2 (\vec{u} \cdot \vec{n}_2 + (\vec{m}_2 \cdot \vec{e}_1)(d - \overrightarrow{Ox} \cdot \vec{n}_2)) \Big]
\end{aligned}$$

9.3.5 La transposée du Jacobien

Nous avons une expression de la dérivée suivant un vecteur des sommets du polygone. Pour calculer le gradient de radiosité, on utilise la transposée du Jacobien.

Il y a deux méthodes pour calculer la transposée du Jacobien d'un sommet :

– nous pouvons exprimer le Jacobien sous forme matricielle et transposer cette forme matricielle. L'application de la transposée du Jacobien à un vecteur se fait alors par une multiplication matrice-vecteur ;

– en utilisant les formes particulières du Jacobien d'un sommet que nous avons obtenu plus haut, et en utilisant la définition même de la transposée d'un opérateur, nous pouvons tirer une expression plus simple de la transposée.

Dans ce dernier cas, le Jacobien est l'opérateur linéaire tel que :

$$\boldsymbol{J}_M : \vec{v} \mapsto \frac{dM}{d\vec{v}}$$

La transposée du Jacobien est donc l'opérateur linéaire tel que :

$$^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto \vec{k}$$

$$\vec{a} \cdot \boldsymbol{J}_M \vec{v} = {}^{\mathrm{t}}\boldsymbol{J}_M \vec{a} \cdot \vec{v} = \vec{k} \cdot \vec{v}$$

La première méthode, l'expression matricielle du Jacobien a pour elle la simplicité d'emploi, d'expression et d'implémentation. En revanche, elle nécessite davantage de calculs.

La deuxième méthode est plus difficile à implémenter, puisqu'il faut distinguer les différents types de sommets au cœur de l'implémentation, mais elle nécessite moins de calculs pour les cas simples, comme les sommets de type 1,2 ou 3.

Type 1

Le Jacobien d'un sommet de type 1 est nul, sa transposée est aussi nulle :

$$^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto \vec{0}$$

Type 2

Le Jacobien d'un sommet de type 2 peut se mettre sous la forme :

$$\boldsymbol{J}_M : \vec{v} \mapsto (\vec{m} \cdot \vec{v})\overrightarrow{EF}$$

Sa transposée s'exprime par :

$$^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto (\vec{a} \cdot \overrightarrow{EF})\vec{m}$$

Ou encore (voyez figure 9.4 pour les notations) :

$$^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto (\vec{a} \cdot \overrightarrow{EF}) \frac{\overrightarrow{PQ} \times \overrightarrow{PF}}{[\overrightarrow{xO} \cdot (\overrightarrow{PQ} \times \overrightarrow{EF})]^2}$$

Type 3

Le Jacobien d'un sommet de type 3 peut se mettre sous la forme :

$$\boldsymbol{J}_M : \vec{v} \mapsto C\vec{n}_2 \times (\vec{v} \times \vec{m})$$

Sa transposée s'exprime par :

$$^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto C\vec{m} \times (\vec{a} \times \vec{n}_2)$$

Ou encore (figure 9.5) :

$$^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto \frac{\overrightarrow{OP} \cdot \vec{n}_2 - d}{(\overrightarrow{xP} \cdot \vec{n}_2)^2} \overrightarrow{xP} \times (\vec{a} \times \vec{n}_2)$$

Type 4

Le Jacobien d'un sommet de type 4 peut se mettre sous la forme :

$$\boldsymbol{J}_M : \vec{v} \mapsto (\vec{m}_1 \cdot \vec{v})\vec{C}_1 + (\vec{m}_2 \cdot \vec{v})\vec{C}_2$$

Sa transposée s'exprime par :

$${}^{\mathrm{t}}\boldsymbol{J}_M : \vec{a} \mapsto (\vec{C}_1 \cdot \vec{a})\vec{m}_1 + (\vec{C}_2 \cdot \vec{a})\vec{m}_2$$

Ou encore (figure 9.6) :

$$\begin{aligned} {}^{\mathrm{t}}\boldsymbol{J}_M \quad : \quad & \vec{a} \mapsto \frac{1}{(\vec{u}\cdot\vec{n}_2)^2}\Big[(\vec{a}\cdot(\vec{n}_2\times\vec{m}_1))\vec{m}_2(\vec{u}\cdot\vec{n}_2 - (\vec{m}_1\cdot\vec{e}_2)(d - \overrightarrow{Ox}\cdot\vec{n}_2)) \\ - \quad & (\vec{a}\cdot(\vec{n}_2\times\vec{m}_2))\vec{m}_1(\vec{u}\cdot\vec{n}_2 + (\vec{m}_2\cdot\vec{e}_1)(d - \overrightarrow{Ox}\cdot\vec{n}_2))\Big] \end{aligned}$$

9.4 Implémentation et résultats

Le principal problème lors de l'implémentation consiste à calculer la portion de l'émetteur qui est visible du récepteur, et les Jacobiens des sommets qui ne sont pas fixes, s'il y a lieu. C'est un problème classique de visibilité.

Étant donné le point x où l'on veut calculer le gradient, on procède de manière incrémentale, en maintenant en permanence une liste d'arêtes correspondant à la partie de l'émetteur qui est encore visible. Pour chaque obstacle, on traite les arêtes de l'obstacle les unes après les autres. Pour chaque arête, on prend le plan P défini par l'arête et par x ; on parcours ensuite la liste des arêtes de l'émetteur, et on cherche les intersections entre les arêtes et le plan.

On voit sur la figure 9.7 la radiosité et la norme du gradient de radiosité dû à un demi-plan qui masque partiellement un émetteur carré. On peut séparer dans les calculs deux parties pour le gradient, d'un côté le terme que l'on a calculé au chapitre précédent, qui est l'intégrale sur le contour de $A_2(x)$ du gradient de $f(x,y)$, et de l'autre le terme calculé ici, qui est une somme de termes dépendant du Jacobien des sommets. Tous deux dépendent de la visibilité, mais de façon différente : le premier ne dépend que de la partie de la source qui est visible, le deuxième dépend aussi des obstacles.

Ces deux composantes, ou plutôt leurs normes, sont visualisés figure 9.8. On voit que le terme qui dépend des obstacles a une influence très localisée. Sur une petite partie du récepteur, il est prépondérant, et négligeable devant le premier terme partout ailleurs.

Si on rapproche l'obstacle de l'émetteur (voir figure 9.9), la variation de radiosité due à l'obstacle devient plus douce, la pente diminue. Si l'on observe les deux termes du gradient (figure 9.10), on voit que le terme dû aux obstacles est devenu négligeable presque partout.

9.5 Calcul du Hessien

Nous avons vu que pour calculer le Hessien, il faut calculer la dérivée seconde suivant deux vecteurs :

$$\frac{\partial^2 B}{\partial\vec{u}\partial\vec{v}}(x)$$

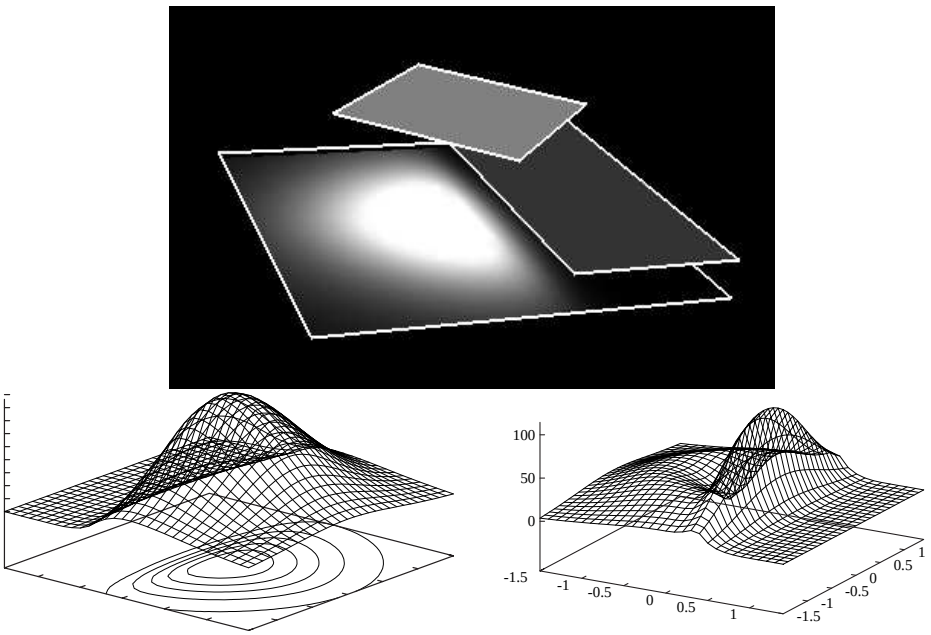
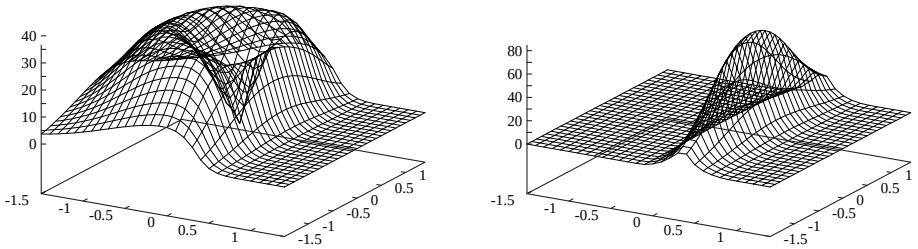


FIG. 9.7 – *La radiosité et la norme du gradient en présence d'obstacles.*



Partie due à la source Partie due aux obstacles

FIG. 9.8 – *Les deux composantes du gradient en présence d'obstacles.*

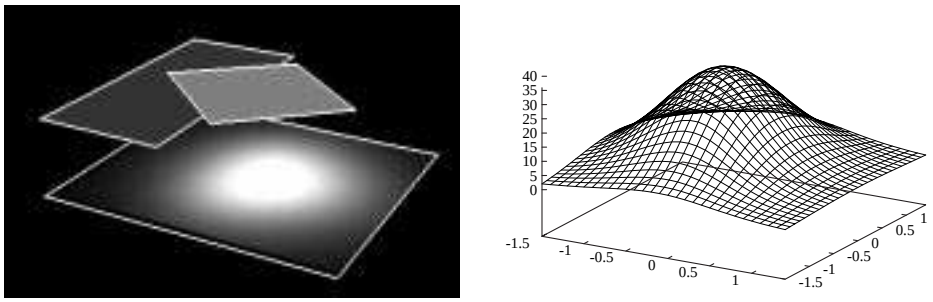
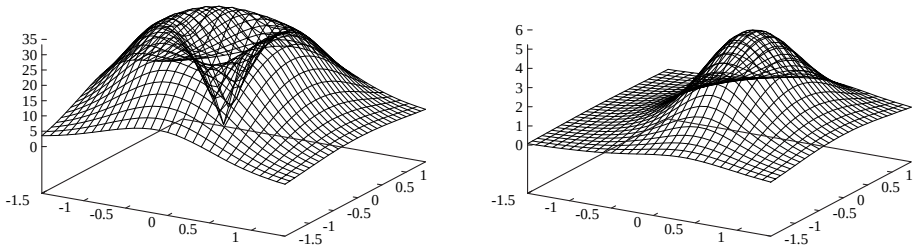


FIG. 9.9 – *Lorsque l'obstacle se situe plus près de la source.*



Partie due à la source Partie due aux obstacles

FIG. 9.10 – *Les deux composantes du gradient en présence d'obstacles.*

puis chercher un opérateur $\boldsymbol{H}$ tel que :

$$\frac{\partial^2 B}{\partial\vec{u}\partial\vec{v}}(x) = {}^{\mathrm{t}}\vec{u}\boldsymbol{H}\vec{v}$$

Comme nous n'avons pas d'expression complète du Jacobien de l'opérateur d'intégration, il nous faut revenir à la définition première de la dérivée seconde, comme limite double :

$$\frac{\partial^2 B}{\partial\vec{u}\partial\vec{v}}(x,y) = \lim_{h\to 0}\lim_{h'\to 0}\frac{1}{h}\frac{1}{h'}\left(\int_{A_2(x+h\vec{v}+h'\vec{u})} f(x+h\vec{v}+h'\vec{u},y)dA_2 - \int_{A_2(x)} f(x,y)dA_2\right)$$

Un travail de réécriture simple, comme celui effectué dans équation 9.1, nous permet de séparer l'expression de la dérivée seconde en plusieurs termes :

$$\begin{aligned}\frac{\partial^2 B}{\partial\vec{u}\partial\vec{v}}(x) &= \int_{A_2(x)} {}^{\mathrm{t}}\vec{u}\boldsymbol{H}_f\vec{v}dA_2 + \vec{v}\cdot\boldsymbol{P}(\vec{u}) + \vec{u}\cdot\boldsymbol{P}(\vec{v}) \\ &+ \lim_{h\to 0}\lim_{h'\to 0}\frac{1}{h}\frac{1}{h'}\left(\int_{A_2(x+h\vec{v}+h'\vec{u})-A_2(x+h\vec{v})-A_2(x+h'\vec{u})+A_2(x)} f(x,y)dA_2\right)\end{aligned}$$

On a noté :

$$\boldsymbol{P}(\vec{u}) = \lim_{h'\to 0}\frac{1}{h'}\int_{A_2(x+h'\vec{u})-A_2(x)} \nabla f(x,y)dA_2$$

Ce terme apparaît deux fois dans l'expression de la dérivée seconde. On peut aussi le voir comme la dérivée par rapport aux changements de visibilité du terme calculé au chapitre 8 pour le gradient de $B(x)$. On le décompose en somme sur les arêtes du polygone :

$$\boldsymbol{P}(\vec{u}) = -\frac{B_0\rho}{2\pi}\sum_i \vec{n}_1\times\frac{\partial}{\partial\vec{u}}(\vec{e}_i I_1) + 2\vec{n}_1\cdot\frac{\partial}{\partial\vec{u}}((\vec{r}_i\times\vec{r}_{i+1})(\vec{r}_i I_2 + \vec{e}_i J_2))$$

On a en fait exprimé par $\frac{\partial}{\partial\vec{u}}$ une dérivée qui ne porte que sur la variation du domaine d'intégration.

$$\frac{\partial}{\partial\vec{u}}(\vec{e}_i I_1) = \delta\boldsymbol{J}_i I_1\vec{u} - I_2(\boldsymbol{J}_i\vec{u}\cdot\vec{r}_i)\vec{e}_i - J_2(\delta\boldsymbol{J}_i\vec{u}\cdot\vec{r}_i + \boldsymbol{J}_i\vec{u}\cdot\vec{e}_i)\vec{e}_i - K_2(\delta\boldsymbol{J}_i\vec{u}\cdot\vec{e}_i)\vec{e}_i$$

Cette dérivée sera utilisée dans une expression de la forme :

$$\vec{v}\cdot\frac{\partial}{\partial\vec{u}}(\vec{e}_i I_1)+\vec{u}\cdot\frac{\partial}{\partial\vec{v}}(\vec{e}_i I_1)={}^{\mathrm{t}}\vec{v}\boldsymbol{H}_{0,i}\vec{u}$$

où $\boldsymbol{H}_{0,i}$ est une matrice symétrique. En utilisant les techniques décrites section 4.7, on peut exprimer $\boldsymbol{H}_{0,i}$:

$$\boldsymbol{H}_{0,i}=I_1(\delta\boldsymbol{J}_i+{}^{\mathrm{t}}\delta\boldsymbol{J}_i)-I_2\boldsymbol{Q}({}^{\mathrm{t}}\boldsymbol{J}_i\vec{r}_i,\vec{e}_i)-J_2\boldsymbol{Q}({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{r}_i+{}^{\mathrm{t}}\boldsymbol{J}_i\vec{e}_i,\vec{e}_i)-K_2\boldsymbol{Q}({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{e}_i,\vec{e}_i)$$

De la même manière, le terme du Hessien dû à :

$$2\vec{v}\cdot\vec{n}_1\cdot\frac{\partial}{\partial\vec{u}}((\vec{r}_i\times\vec{r}_{i+1})(\vec{r}_i I_2+\vec{e}_i J_2))+2\vec{u}\cdot\vec{n}_1\cdot\frac{\partial}{\partial\vec{v}}((\vec{r}_i\times\vec{r}_{i+1})(\vec{r}_i I_2+\vec{e}_i J_2))={}^{\mathrm{t}}\vec{v}\boldsymbol{H}_{1,i}\vec{u}$$

se calcule et donne :

$$\begin{aligned}
\boldsymbol{H}_{1,i} &= \boldsymbol{Q}({}^{\mathrm{t}}\boldsymbol{J}_{i+1}(\vec{n}_1\times\vec{r}_i)-{}^{\mathrm{t}}\boldsymbol{J}_i(\vec{n}_1\times\vec{r}_{i+1}),\vec{r}_i I_2+\vec{e}_i J_2)\\
&+ \vec{n}_1\cdot(\vec{r}_i\times\vec{r}_{i+1})(I_2(\boldsymbol{J}_i+{}^{\mathrm{t}}\boldsymbol{J}_i)+J_2(\delta\boldsymbol{J}_i+{}^{\mathrm{t}}\delta\boldsymbol{J}_i))\\
&- 4\vec{n}_1\cdot(\vec{r}_i\times\vec{r}_{i+1})(\boldsymbol{Q}((I_3{}^{\mathrm{t}}\boldsymbol{J}_i+J_3{}^{\mathrm{t}}\delta\boldsymbol{J}_i)\vec{r}_i,\vec{r}_i)+\boldsymbol{Q}((J_3{}^{\mathrm{t}}\boldsymbol{J}_i+K_3\delta\boldsymbol{J}_i)\vec{r}_i,\vec{e}_i))\\
&- 4\vec{n}_1\cdot(\vec{r}_i\times\vec{r}_{i+1})(+\boldsymbol{Q}((J_3{}^{\mathrm{t}}\boldsymbol{J}_i+K_3\delta\boldsymbol{J}_i)\vec{e}_i,\vec{r}_i)+\boldsymbol{Q}((K_3{}^{\mathrm{t}}\boldsymbol{J}_i+L_3\delta\boldsymbol{J}_i)\vec{e}_i,\vec{e}_i))
\end{aligned}$$

Avec $\boldsymbol{H}_{0,i}$ et $\boldsymbol{H}_{1,i}$, on a calculé la partie du Hessien qui est dû aux variations croisées : la composée d'une dérivation avec variation du contour (la fonction intégrée restant fixe) et d'une dérivation de la fonction intégrée, le contour restant fixe.

Le terme qui nous reste à calculer est le terme correspondant à la dérivée seconde du contour :

$${}^{\mathrm{t}}\vec{v}H_3\vec{u}=\lim_{h\to0}\lim_{h'\to0}\frac{1}{h}\frac{1}{h'}\left(\int_{A_2(x+h\vec{v}+h'\vec{u})-A_2(x+h\vec{v})-A_2(x+h'\vec{u})+A_2(x)}f(x,y)dA_2\right)$$

On peut aussi exprimer H_3 comme une intégrale de contours :

$${}^{\mathrm{t}}\vec{v}H_3\vec{u}=\vec{n}_1\cdot\lim_{h\to0}\lim_{h'\to0}\frac{1}{h}\frac{1}{h'}\left(\int_{\partial A_2(x+h\vec{v}+h'\vec{u})-\partial A_2(x+h\vec{v})-\partial A_2(x+h'\vec{u})+\partial A_2(x)}\frac{\vec{r}_{12}}{\|\vec{r}_{12}\|^2}\times d\vec{\ell}_2\right)\tag{9.4}$$

Pour l'instant, supposons que nous sachions calculer le Hessien d'un sommet ; dans ce cas, on a (en reprenant les notations de la figure 9.1) :

$$E'i=E_i+h\boldsymbol{J}_i\vec{v}+h^2{}^{\mathrm{t}}\vec{v}\boldsymbol{H}_i\vec{v}+\vec{o}(h^2)$$

Si on effectue un développement limité de l'intégrale de l'équation 9.4, les termes en h, h^2, h^3,... s'éliminent, de même que les termes en h', h'^2, h'^3,..., et de même que le terme constant. D'autre part, les termes en $h^n h'^{n'}$ avec n et n' strictement supérieurs à un ont une limite nulle après division par hh'. Il ne reste donc que le terme en hh' à calculer dans le développement limité de l'intégrale.

$$\begin{aligned}
{}^{\mathrm{t}}\vec{v}\boldsymbol{H}_3\vec{u} &= \vec{n}_1\cdot\sum_i{}^{\mathrm{t}}\vec{v}\boldsymbol{H}_{3,i}\vec{u}\\
{}^{\mathrm{t}}\vec{v}\boldsymbol{H}_{3,i}\vec{u} &= -\int_0^1\frac{{}^{\mathrm{t}}\vec{v}(\boldsymbol{H}_i+y\delta H_i)\vec{u}\cdot(\vec{r}_i+y\vec{e}_i)}{\|\vec{r}_{12}\|^4}dy\\
&+ 2\int_0^1\frac{(\boldsymbol{J}_i\vec{u}+y\delta\boldsymbol{J}_i\vec{u})\cdot(\vec{r}_i+y\vec{e}_i)(\boldsymbol{J}_i\vec{v}+y\delta\boldsymbol{J}_i\vec{v})\cdot(\vec{r}_i+y\vec{e}_i)}{\|\vec{r}_{12}\|^6}dy
\end{aligned}$$

Le terme où apparaît le Hessien des sommets est un terme ou le Hessien opère sur deux vecteurs ($\vec{u}$ et $\vec{v}$), ce qui donne un vecteur (le Hessien d'un sommet est un opérateur 3×3). On prend ensuite le produit scalaire de ce vecteur avec $\vec{r}_i$ et $\vec{e}_i$. Au prix d'une réorganisation de $\boldsymbol{H}_i$, on peut écrire l'opération de $\boldsymbol{H}_i$ sur les deux vecteurs $\vec{u}$ et $\vec{v}$ comme : $(\boldsymbol{H}_i\vec{u})\vec{v}$. Auquel cas,

$$((\boldsymbol{H}_i\vec{u})\vec{v})\cdot\vec{r}_i = {}^{\mathrm{t}}\vec{r}_i(\boldsymbol{H}_i\vec{u})\vec{v} = ({}^{\mathrm{t}}({}^{\mathrm{t}}\boldsymbol{H}_i\vec{r}_i)\vec{u})\vec{v}$$

Nous verrons plus loin quelle forme prend le Hessien d'un sommet, et surtout sa transposée.

Avec ces notations,

$$\begin{aligned}
\boldsymbol{H}_{3,i} &= -\int_0^1 \frac{({}^{\mathrm{t}}\boldsymbol{H}_i + y\,{}^{\mathrm{t}}\delta H_i)(\vec{r}_i+y\vec{e}_i)}{\|\vec{r}_{12}\|^4}dy \\
&+ 2\int_0^1 \frac{\boldsymbol{Q}(({}^{\mathrm{t}}\boldsymbol{J}_i+y\delta\,{}^{\mathrm{t}}\boldsymbol{J}_i)(\vec{r}_i+y\vec{e}_i),({}^{\mathrm{t}}\boldsymbol{J}_i+y\delta\,{}^{\mathrm{t}}\boldsymbol{J}_i)(\vec{r}_i+y\vec{e}_i)(\vec{r}_i+y\vec{e}_i))}{\|\vec{r}_{12}\|^6}dy \\
\boldsymbol{H}_{3,i} &= -{}^{\mathrm{t}}\boldsymbol{H}_i(\vec{r}_iI_2+\vec{e}_iJ_2) - {}^{\mathrm{t}}\delta\boldsymbol{H}_i(\vec{r}_iJ_2+\vec{e}_iK_2) \\
&+ \boldsymbol{Q}({}^{\mathrm{t}}\boldsymbol{J}_i\vec{r}_i, I_3\,{}^{\mathrm{t}}\boldsymbol{J}_i\vec{r}_i + J_3({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{r}_i + {}^{\mathrm{t}}\boldsymbol{J}_i\vec{e}_i) + K_3\,{}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{e}_i) \\
&+ \boldsymbol{Q}({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{r}_i + {}^{\mathrm{t}}\boldsymbol{J}_i\vec{e}_i, J_3\,{}^{\mathrm{t}}\boldsymbol{J}_i\vec{r}_i + K_3({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{r}_i + {}^{\mathrm{t}}\boldsymbol{J}_i\vec{e}_i) + L_3\,{}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{e}_i) \\
&+ \boldsymbol{Q}({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{e}_i, K_3\,{}^{\mathrm{t}}\boldsymbol{J}_i\vec{r}_i + L_3({}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{r}_i + {}^{\mathrm{t}}\boldsymbol{J}_i\vec{e}_i) + M_3\,{}^{\mathrm{t}}\delta\boldsymbol{J}_i\vec{e}_i)
\end{aligned}$$

On voit que le Hessien est calculable, bien que complexe, et qu'il se décompose en trois termes : le premier est le terme que nous avions au chapitre 8, une intégrale sur l'émetteur, et correspond aux variations de la fonction intégrée sur l'émetteur, en ne tenant pas compte des variations de surface de l'émetteur, le dernier correspond à la dérivée seconde des sommets de l'émetteur, à cause des obstacles entre l'émetteur et le récepteur, et le terme central correspond aux effets conjoints de la dérivée de la radiosité sur l'émetteur et de la dérivée de l'émetteur à cause des obstacles.

Contrairement au gradient de radiosité, qui dépendait presque exclusivement du Jacobien des sommets, parmi les nombreux termes qui composent le Hessien de la radiosité, un seul nécessite le calcul du Hessien des sommets. On pourrait donc envisager de calculer une valeur approchée du Hessien de la radiosité en utilisant uniquement les termes indépendants du Hessien des sommets.

Pour calculer de façon complète le Hessien de la radiosité en présence d'obstacles, il faudrait calculer le Hessien des sommets de l'émetteur. Un sommet ayant trois paramètres (x,y,z), son Jacobien est une matrice 3×3, et son Hessien est un opérateur $3\times 3\times 3$. Plutôt que l'expression du Hessien, nous devrons rechercher l'expression de l'opérateur $\boldsymbol{M}$ tel que $({}^{\mathrm{t}}\vec{v}\boldsymbol{H}_i\vec{u})\cdot\vec{r}_i = {}^{\mathrm{t}}\vec{v}(\boldsymbol{M}\vec{r}_i)\vec{u}$.

10.

Extension au cas des émetteurs non constants

NOUS avons vu aux chapitres précédents comment une simple manipulation de l'équation de radiosité donnait accès à des informations supplémentaires, et notamment à une expression exacte du gradient de la radiosité.

Ce calcul de l'expression du gradient reposait en grande partie sur l'existence d'une expression de la radiosité sous forme d'une intégrale de contour. Nous avons vu au chapitre 4 que lorsque la radiosité sur l'émetteur n'est pas constante, il est impossible d'avoir une expression de la radiosité sur le récepteur sous forme d'une intégrale de contour. En revanche, on a une expression de la radiosité sous forme d'une somme, d'une part une intégrale de contour, et d'autre part une intégrale surfacique :

$$B(x) = E(x) - \vec{n}_1 \cdot \frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{\vec{r}_{12} \times d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} - \vec{n}_1 \cdot \frac{\rho}{2\pi} \int_{A_2} \left(\frac{\vec{k}(y) \times \vec{r}_{12}}{\|\vec{r}_{12}\|^2} \right) dA_2$$

Où on a noté $\vec{k}(y) = \vec{n}_2 \times \nabla B(y)$. Cette expression de la radiosité permet parfois de calculer la radiosité et son gradient de façon plus précise qu'avec la seule intégrale surfacique.

10.1 Calcul de la radiosité

Le calcul de l'intégrale surfacique est un point délicat. On peut remarquer que le terme surfacique :

$$T_S = -\vec{n}_1 \cdot \frac{\rho}{2\pi} \int_{A_2} \left(\frac{\vec{k}(y) \times \vec{r}_{12}}{\|\vec{r}_{12}\|^2} \right) dA_2$$

peut s'exprimer en fait comme :

$$T_S = -\vec{n}_1 \cdot \frac{\rho}{2\pi} \int_{A_2} \vec{k}(y) \times \nabla \ln(r_{12}) dA_2$$

Nous ne pouvons pas aller plus loin sans introduire des hypothèses sur le comportement de B et, partant, de $\vec{k}$. Ces hypothèses dépendent en fait de la façon dont on représente la radiosité sur l'émetteur ; si on choisit une base de fonctions, on décompose B sur cette base de fonctions, $\vec{k}$ est ainsi décomposé en somme de composantes, et l'on calcule T_S pour chacune de ces composantes.

Par exemple, si on suppose une radiosité qui varie de façon linéaire sur l'émetteur, ∇B est constant, et $\vec{k}$ aussi. Dans ce cas, il ne nous reste qu'à calculer :

$$\vec{m} = \int_{A_2} \nabla \ln(r_{12}) dA_2 = \int_{A_2} \frac{\vec{r}_{12}}{\|r_{12}\|^2} dA_2$$

On peut exprimer directement $\vec{m}$ en utilisant une approximation de l'intégrale vectorielle. On peut aussi remarquer que $\vec{m} \times \vec{n}_2$ peut être transformé en une intégrale de contours en utilisant le théorème d'Ostrogradsky (voir équation 4.2) :

$$\vec{m} \times \vec{n}_2 = \int_{A_2} \nabla \ln(r_{12}) \times d\vec{A}_2 = \int_{\partial A_2} \ln(r_{12}) d\vec{\ell}_2$$

Dans ce cas, il ne nous reste plus qu'à calculer $\vec{m} \cdot \vec{n}_2$. Si nous sommes dans un cas où $\vec{m}$ n'est pas nul, alors la distance r_{12} ne s'annule jamais sur A_2. Par ailleurs, $\vec{r}_{12} \cdot \vec{n}_2$ ne varie par lorsque le point courant décrit A_2. On peut choisir un événement $\vec{r}_O$ quelconque,

$$\vec{m} \cdot \vec{n}_2 = \vec{r}_0 \cdot \vec{n}_2 \int_{A_2} \frac{dA_2}{\|r_{12}\|^2} = \vec{r}_0 \cdot \vec{n}_2 X_1$$

La fonction dont on doit calculer l'intégrale surfacique est $\frac{1}{\|r_{12}\|^2}$; on connaît les bornes du domaine d'intégration, et on peut calculer la distance minimale entre x et l'émetteur, ainsi que la distance maximale. On a alors un minorant et un majorant de l'intégrale.

On pourrait aussi calculer une approximation plus précise de l'intégrale, en utilisant les différentes méthodes d'approximation des intégrales (méthode des rectangles, méthode de Simpson, méthode de Romberg...). L'important est de conserver toujours un encadrement de l'intégrale et non une simple valeur approchée, afin de pouvoir juger de la précision sur la radiosité au cours des calculs.

Une fois qu'on connaît $\vec{m} \cdot \vec{n}_2$ et $\vec{m} \times \vec{n}_2$, on peut calculer $\vec{m}$:

$$\vec{m} = (\vec{m} \cdot \vec{n}_2)\vec{n}_2 + \vec{n}_2 \times (\vec{m} \times \vec{n}_2)$$

D'où l'expression de la radiosité due à un émetteur sur lequel la radiosité varie de façon linéaire :

$$B(x) = E(x) - \vec{n}_1 \cdot \frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{\vec{r}_{12} \times d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} - \frac{\rho}{2\pi} \vec{n}_1 \cdot (\vec{k} \times \vec{n}_2)(\vec{m} \cdot \vec{n}_2) - \frac{\rho}{2\pi} (\vec{n}_1 \cdot \vec{n}_2)(\vec{k} \cdot (\vec{m} \times \vec{n}_2))$$

(dans cette expression, on a utilisé le fait que $\vec{k} \cdot \vec{n}_2 = 0$). Si l'on exprime $\vec{m} \times \vec{n}_2$ comme une intégrale de contour, il vient :

$$\begin{aligned} B(x) &= E(x) - \vec{n}_1 \cdot \frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{\vec{r}_{12} \times d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} \\ &- \frac{\rho}{2\pi} \vec{n}_1 \cdot (\vec{k} \times \vec{n}_2)(\vec{r}_0 \cdot \vec{n}_2) X_1 - \frac{\rho}{2\pi} (\vec{n}_1 \cdot \vec{n}_2)(\vec{k} \cdot \int_{\partial A_2} \ln(r_{12}) d\vec{\ell}_2) \end{aligned}$$

L'encadrement sur la valeur de X_1 donne, non pas une valeur exacte pour $B(x)$, mais un encadrement :

$$B_{sup} > B(x) > B_{inf}$$

On remarque que l'incertitude sur X_1 a été multipliée par $\vec{n}_1 \cdot (\vec{k} \times / n_2) = \vec{n}_1 \cdot \nabla B(y)$, qui est inférieur à la norme de $\nabla B(y)$. On retrouve ainsi que l'on a une valeur exacte lorsque la radiosité est constante sur l'émetteur.

On remarque aussi que l'incertitude est annulée lorsque $\vec{n}_1 \cdot \nabla B(y) = 0$. Ceci est en particulier vérifié si $\vec{n}_1$ et $\vec{n}_2$ sont colinéaires. Donc si l'émetteur et le récepteur sont parallèles, on a à nouveau une valeur exacte. $\vec{n}_1 \cdot \nabla B(y)$ est aussi vérifié dans certains cas, si le plan récepteur est judicieusement disposé par rapport à l'émetteur.

Comme $\nabla B(y)$ et $\vec{n}_2$ sont perpendiculaires, on peut les utiliser comme base pour bâtir un repère orthonormé direct : $(\frac{\nabla B}{\|\nabla B\|}, \frac{\vec{k}}{\|\vec{k}\|}, \vec{n}_2)$. On peut ensuite exprimer la position de $\vec{n}_1$ dans ce repère, puis la convertir en coordonnées sphériques (θ, ϕ). L'incertitude sur X_1 est alors multipliée par $\sin\theta\cos\phi\|\nabla B(y)\|$. On peut calculer la moyenne de ce terme multiplicateur : $\frac{1}{2}\|\nabla B(y)\|$. En moyenne, l'erreur sur X_1 sera donc multipliée par $\frac{1}{2}\|\nabla B(y)\|$. On constate que l'utilisation du théorème d'Ostrogradsky a permis de réduire l'erreur sur la radiosité, et de la localiser dans une seule direction, ce que l'on n'aurait pas obtenu avec une approximation directe de $\vec{m}$. Dans ce contexte, la formule introduite au chapitre 4 a bel et bien permis d'augmenter la précision par rapport à l'approximation numérique directe sur l'intégrale surfacique.

En particulier, le terme multiplicatif $T = \vec{n}_1(\vec{k} \times \vec{n}_2)(\vec{r}_0 \cdot \vec{n}_2)$ peut être calculé avant l'approximation de l'intégrale X_1. Si l'on sait que l'on veut modéliser la radiosité à ε près, il suffit de calculer X_1 à $\frac{\varepsilon}{T}$ près.

10.2 Calcul du gradient

Dans l'expression de B, ci-dessus, le premier terme est exactement celui que nous avions au chapitre 8, la seule différence étant que la radiosité de l'émetteur a cessé d'être supposée constante, et donc a été passée sous le signe d'intégration. Comme $B(y)$ ne dépend pas de la position du point x, la dérivation formelle de ce terme se conduit de la même manière qu'au chapitre 8. Le deuxième terme contient une intégrale sur un contour, qui se dérive également sans peine. Le terme en $\vec{m} \cdot \vec{n}_2$ introduit quelques difficultés :

$$\nabla\left(\frac{\vec{r}_{12} \cdot \vec{n}_2}{\|r_{12}\|^2}\right) = \frac{\vec{n}_2}{\|r_{12}\|^2} - 2\vec{r}_{12}\frac{\vec{r}_{12} \cdot \vec{n}_2}{\|r_{12}\|^4}$$

D'où :

$$\nabla(\vec{m} \cdot \vec{n}_2) = \nabla\left(\int_{A_2} \frac{\vec{r}_{12} \cdot \vec{n}_2}{\|r_{12}\|^2} dA_2\right) = \vec{n}_2 \int_{A_2} \frac{dA_2}{\|r_{12}\|^2} + (\vec{n}_2 \cdot \vec{r}_0)\vec{p}$$

On a noté :

$$\vec{p} = \int_{A_2} -\frac{2\vec{r}_{12}}{\|r_{12}\|^4} dA_2 = \int_{A_2} \nabla\left(\frac{1}{\|r_{12}\|^2}\right) dA_2$$

On remarque que $\vec{p} \times \vec{n}_2$ se calcule aisément, en utilisant à nouveau le théorème d'Ostrogradsky :

$$\vec{p} \times \vec{n}_2 = \int_{\partial A_2} \frac{d\vec{\ell}_2}{\|r_{12}\|^2}$$

Pour calculer $\vec{p} \cdot \vec{n}_2$, on est à nouveau obligé de requérir à une approximation :

$$\vec{p} \cdot \vec{n}_2 = (\vec{n}_2 \cdot \vec{r}_0) \int_{A_2} \frac{1}{\|r_{12}\|^4} dA_2 = (\vec{n}_2 \cdot \vec{r}_0) X_2$$

On a donc l'expression du gradient dans le cas particulier où on s'est placé :

$$\begin{aligned}
\nabla B(x) &= \nabla E(x) + \vec{n}_1 \times \frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{d\vec{\ell}_2}{\|\vec{r}_{12}\|^2} \\
&+ 2\frac{\rho}{2\pi} \oint_{\partial A_2} B(y) \frac{\vec{n}_1 \cdot \vec{r}_{12}}{\|r_{12}\|^4} (\vec{r}_{12} \times d\vec{\ell}_2) \\
&- \frac{\rho}{2\pi} (\vec{n}_1 \cdot \vec{n}_2) \int_{\partial A_2} \frac{\vec{r}_{12}}{\|r_{12}\|^2} (\vec{k} \cdot d\vec{\ell}_2) \\
&- \frac{\rho}{2\pi} \vec{n}_1 \cdot (\vec{k} \times \vec{n}_2) \left(X_1 \vec{n}_2 + (\vec{r}_0 \cdot \vec{n}_2)^2 X_2 \vec{n}_2 + (\vec{r}_0 \cdot \vec{n}_2) \vec{n}_2 \times \oint_{\partial A_2} \frac{d\vec{\ell}_2}{\|r_{12}\|^2} \right)
\end{aligned}$$

10.3 Implémentation

On reprend les notations de 8.4, en paramétrant chaque arête par $y \in [0,1]$:

$$\begin{aligned}
\vec{r}_{12} &= \vec{r}_i + y\vec{e}_i \\
d\vec{\ell}_2 &= \vec{e}_i dy \\
\vec{r}_{12} \times d\vec{\ell}_2 &= (\vec{r}_i \times \vec{e}_i) dy \\
&= (\vec{r}_i \times \vec{r}_{i+1}) dy \\
B(y) &= B_i + y\delta B_i \\
\delta B_i &= B_{i+1} - B_i
\end{aligned}$$

De la même façon qu'aux chapitres 8 et 9, on introduit les quantités I_n, J_n et K_n :

$$\begin{aligned}
I_n &= \int_0^1 \frac{dy}{\|\vec{r}_{12}\|^{2n}} \\
J_n &= \int_0^1 \frac{y dy}{\|\vec{r}_{12}\|^{2n}} \\
K_n &= \int_0^1 \frac{y^2 dy}{\|\vec{r}_{12}\|^{2n}}
\end{aligned}$$

On connaît les valeurs de I_n, J_n et K_n (voir section 8.4 et section 8.7). On introduit par surcroît la valeur L_1 :

$$L_1 = \int_0^1 \ln(r_{12}) dy$$

On a :

$$L_1 = \frac{1}{e_i^2} \left((\vec{r}_{i+1} \cdot \vec{e}_i) \ln(e_{i+1}) - (\vec{r}_i \cdot \vec{e}_i) \ln(e_i) + 2\|\vec{r}_i \times \vec{e}_i\| \gamma_i \right)$$

On obtient l'expression de B :

$$\begin{aligned}
B(x) &= E(x) - \frac{\rho}{2\pi} \sum_i \vec{n}_1 \cdot (\vec{r}_i \times \vec{e}_i) (B_i I_1 + \delta B_i J_1) \\
&- \frac{\rho}{2\pi} \vec{n}_1 \cdot (\vec{k} \times \vec{n}_2) (\vec{r}_0 \cdot \vec{n}_2) X_1 \\
&- \frac{\rho}{2\pi} (\vec{n}_1 \cdot \vec{n}_2) (\vec{k} \cdot \sum_i L_1 \vec{e}_i)
\end{aligned}$$

Cette expression s'implémente de la même manière qu'aux chapitres 8 et 9. La figure 10.2 montre un exemple de radiosité sur le récepteur due à un émetteur sur lequel la radiosité varie de façon linéaire (voir figure 10.1). La figure 10.3 montre la norme du gradient de radiosité due à cet émetteur. On remarque que la norme du gradient s'annule en un point, qui correspond au maximum de la radiosité sur l'émetteur.

Le calcul pratique du gradient se conduit de la même manière :

$$\begin{aligned}
\nabla B(x) &= \nabla E(x) + \vec{n}_1 \times \frac{\rho}{2\pi} \sum_i (B_i I_1 + \delta B_i J_1) \vec{e}_i \\
&+ 2\frac{\rho}{2\pi} \sum_i (\vec{r}_i \times \vec{e}_i) \left(B_i (\vec{n}_1 \cdot \vec{r}_i) I_2 + (B_i (\vec{n}_1 \cdot \vec{e}_i) + \delta B_i (\vec{n}_1 \cdot \vec{r}_i)) J_2 + \delta B_i (\vec{n}_1 \cdot \vec{e}_i) K_2 \right) \\
&- \frac{\rho}{2\pi} (\vec{n}_1 \cdot \vec{n}_2) \sum_i (\vec{k} \cdot \vec{e}_i) (\vec{r}_i I_1 + \vec{e}_i J_1) \\
&- \frac{\rho}{2\pi} \vec{n}_1 \cdot (\vec{k} \times \vec{n}_2) \left(X_1 \vec{n}_2 + (\vec{r}_0 \cdot \vec{n}_2)^2 X_2 \vec{n}_2 + (\vec{r}_0 \cdot \vec{n}_2) \vec{n}_2 \times \sum_i \vec{e}_i I_1 \right)
\end{aligned}$$

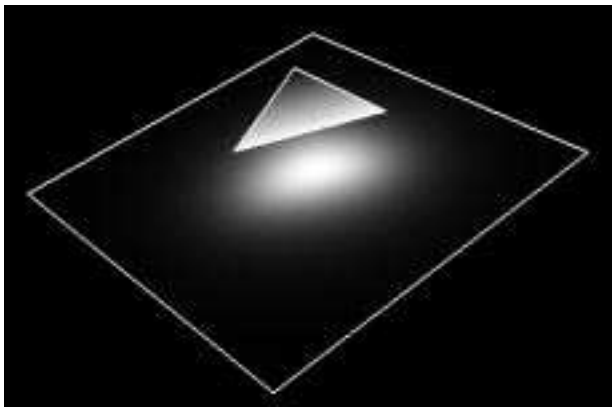


FIG. 10.1 – *Une scène de tests pour un émetteur sur lequel la radiosité n'est pas constante.*

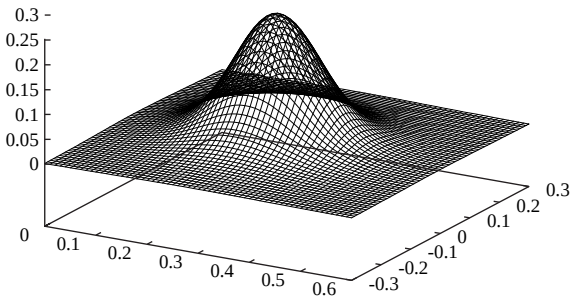


FIG. 10.2 – *La radiosité due à cet émetteur.*

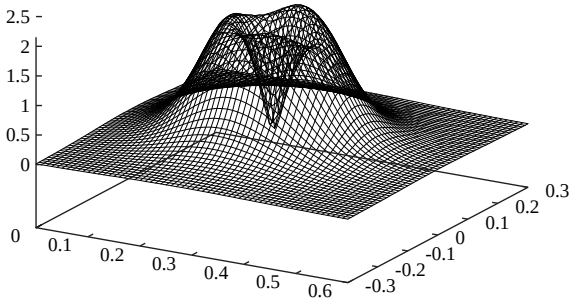


FIG. 10.3 – *Le gradient de la radiosité due à cet émetteur.*

11.

Conclusion et futures directions

Le contrôle de l'erreur au sein des algorithmes de simulation physique – par exemple des algorithmes de simulation physique de la lumière – est l'un des défis les plus prometteurs des années à venir. La connaissance de l'erreur commise dans la modélisation d'un processus physique permet à la fois de valider l'algorithme de résolution et de réduire le coût de la résolution numérique en concentrant les ressources du système sur les points où l'erreur est importante.

Nous avons présenté un critère de raffinement pour un algorithme de radiosité permettant un contrôle total de l'erreur commise sur chaque interaction, ainsi qu'un début de contrôle de l'erreur de discrétisation. Ce critère de raffinement est intégré au sein d'un algorithme de radiosité hiérarchique, permettant ainsi de ne modéliser les interactions que jusqu'à la précision souhaitée, économisant ainsi les ressources système (temps processeur et mémoire). Cet algorithme de contrôle de l'erreur est un outil précieux pour toute application industrielle de la méthode de radiosité hiérarchique. Ce critère de raffinement pourrait également être intégré au sein d'un algorithme de radiosité progressive (tel celui décrit par Cohen *et al.*, 1988 [6]), puisqu'il ne dépend pas de la structure interne du programme de résolution.

Nous avons également présenté un critère de raffinement, dit «simplifié», qui, tout en permettant le même contrôle de l'erreur commise sur chaque interaction que le critère complet, permet de réduire de façon significative le nombre de liens nécessaires à la modélisation de l'interaction. Cette réduction du nombre de liens permet une économie supplémentaire des ressources du système, à la fois en temps et en mémoire.

Ces deux critères de raffinement s'appuient d'une part sur un calcul exact des dérivées de la radiosité, provenant d'une analyse de l'expression de la radiosité, et d'autre part sur des conjectures de concavité, qui sont basées sur une analyse du comportement de la radiosité due à un émetteur convexe.

Notre algorithme repose fortement sur ces deux conjectures de concavité, qui pour l'instant encore ne sont pas démontrées, mais pour lesquelles on ne connait pas de situations où elles soient prises en défaut. Ces conjectures représentent un résumé des propriétés de la radiosité qui sont nécessaires au contrôle de l'erreur. Une démonstration des conjectures de concavité permettrait d'une part de mieux comprendre le comportement de la radiosité, et d'autre part de mieux la modéliser.

Ces propriétés de concavité de la radiosité ne sont pas maintenues en présence d'obstacles, même si ces obstacles sont convexes. Notre algorithme de contrôle de l'erreur repose donc sur le calcul d'un émetteur minimal et d'un émetteur maximal pour chaque interaction, tous deux convexes et encadrant la partie de l'émetteur qui est effectivement visible de chacun des points du récepteur. Une extension des conjectures de concavité au cas d'un émetteur convexe en présence d'obstacles, conjointement avec une implémentation du Hessien en présence d'obstacles, permettrait de réduire les calculs nécessaires pour le contrôle de l'erreur en présence d'obstacles. Dans cette optique, il faudrait calculer non plus l'émetteur minimal et l'émetteur maximal, mais l'obstacle maximal et l'obstacle minimal, encadrant les obstacles réels, et chacun d'eux étant tel que la conjecture de concavité est respectée et que l'on sait calculer le Hessien.

L'obstacle maximal et l'obstacle minimal, pour que la conjecture de concavité puisse être respectée, doivent être de forme très simple. Remarquons que lorsqu'une seule arête de l'obstacle vient bloquer la visibilité de l'émetteur depuis le récepteur, l'obstacle minimal et l'obstacle maximal coïncident. Dans ce cas, on peut espérer une réduction substantielle du nombre de facettes nécessaires, surtout en utilisant le critère «simplifié». Remarquons aussi qu'il est plus facile de calculer le Hessien des sommets lorsqu'il n'y a qu'un seul obstacle de forme demi-plan.

Nous avons donné dans les chapitres précédents une expression des dérivées de la radiosité. Cette expression est directement implémentable, par un simple travail de traduction. En revanche, sa complexité devient quasiment prohibitive dans le cas du Hessien en présence d'obstacles. Les travaux futurs devront porter sur une simplification de notre expression des dérivées successives de la radiosité, permettant ainsi une implémentation effective du Hessien de la radiosité en présence d'obstacles.

Cette expression des dérivées successives de la radiosité (gradient et Hessien) pourrait aussi être utilisée dans une implémentation de la radiosité utilisant des fonctions de base d'ordre supérieur (voir Troutman [43]), où on calcule la radiosité sur les facettes par une interpolation polynomiale à partir des valeurs de la radiosité et de ses dérivées calculées aux sommets des facettes.

Nous avons également présenté une méthode de réécriture de l'expression de la radiosité en un point sous l'influence d'un émetteur sur lequel la radiosité est quelconque. L'expression ainsi obtenue permet un encadrement précis de la radiosité en ce point. Une application au cas où la radiosité sur l'émetteur varie de façon linéaire a été introduite. On a dans ce cas une expression de la radiosité et de son gradient en tout point de l'émetteur, sous forme d'une somme de deux termes, l'un qui est exact, et l'autre pour lequel on dispose d'un encadrement. On a ainsi presque tous les outils nécessaires à un algorithme de radiosité linéaire avec un contrôle complet de l'erreur commise sur chaque interaction. Pour pouvoir implémenter complètement cet algorithme de radiosité linéaire, il est encore nécessaire, d'abord d'obtenir une expression du Hessien dans le cas de la radiosité linéaire, et ensuite d'adapter les conjectures de concavité à ce cas précis – en particulier, il est possible que dans certaines configurations les conjectures de concavité ne soient pas vérifiées. On ne pourra donc utiliser l'hypothèse de la radiosité linéaire sur l'émetteur que dans certains cas précis. Néanmoins, lorsqu'on pourra l'utiliser, on peut s'attendre à une réduction substantielle du nombre de facettes et de liens nécessaires pour modéliser les interactions.

Notre algorithme de contrôle de l'erreur en présence d'obstacles repose sur le calcul de l'émetteur minimal et de l'émetteur maximal, qui dépendent tous deux de la facette réceptrice. Tel qu'il a été présenté, l'algorithme est indépendant de la structure du maillage utilisé sur les objets composant la scène, et il fonctionne même lorsque ce maillage est un simple arbre quadtree. Cependant, il serait intéressant d'intégrer notre algorithme de contrôle de l'erreur au sein d'un algorithme de maillage de discontinuité, tel que la *backprojection* décrite par Drettakis [12, 10]. Dans ce cadre, le calcul des émetteurs minimaux et maximaux serait simplifié au sein de chaque cellule du maillage, puisqu'on a une connaissance précise des arêtes des obstacles qui influent sur la visibilité de l'émetteur. Cette intégration peut se faire de façon simple *a posteriori*, l'algorithme de contrôle de l'erreur sur chaque interaction étant utilisé pour calculer la validité de chacune des facettes créées, et pour les raffiner si besoin est. Mais on peut aussi imaginer une intégration de l'algorithme de contrôle de l'erreur au sein même de l'algorithme de *backprojection*, qui permettrait d'indiquer si le niveau du maillage déjà obtenu est suffisant au regard de la précision souhaitée, ou s'il est nécessaire de poursuivre les calculs.

Nous avons également présenté une analyse complète de l'algorithme de radiosité hiérarchique, qui a permis de souligner les points les plus délicats de cet algorithme et ses goulots d'étranglement, confirmant par exemple que les calculs de visibilité constituent jusqu'à 80 % du coût total de l'algorithme. Cette analyse est en grande partie liée à une implémentation spécifique de la méthode de radiosité hiérarchique, donc à une méthode spécifique de calcul de la visibilité. Une telle analyse demande à être conduite de façon plus systématique : sur un plus grand nombre de scènes, de complexité plus grande, et avec différentes implémentations de la méthode de radiosité hiérarchique. On pourra ainsi quantifier l'intérêt des choix d'implémentation, à la fois en termes de rapidité et de précision. Le jeu de scènes employé pour cette analyse pourrait être diffusé de façon internationale, ce qui permettrait de conduire une série de tests à grande échelle, et une comparaison des différents algorithmes de rendu et de modélisation de la lumière en fonction de la complexité de la scène.

En conclusion, nous avons montré qu'il est possible de contrôler l'erreur commise par l'algorithme de radiosité hiérarchique dans la modélisation des interactions entre les objets. Grâce à ce contrôle, il est possible de garantir les résultats de l'algorithme de modélisation de l'éclairage. La fiabilité ainsi obtenue dans la résolution se fait pour l'instant au détriment de la rapidité de la résolution. Par ailleurs, le contrôle de l'erreur se fait pour l'instant au niveau des interactions entre les objets, ce qui peut être insuffisant lorsque la scène de départ est déjà trop raffinée. La recherche d'un contrôle complet de l'erreur commise au cours de l'ensemble du processus de modélisation, ainsi que l'optimisation des méthodes de contrôle de l'erreur afin de réduire les surcoûts en terme de temps de calcul sont des pistes de recherche fondamentales pour les années à venir.

Bibliographie

[1] APPEL (A.), «Some Techniques for Shading Machines Rendering of Solids», dans *Proceedings of the Spring Joint Computer Conference*, p. 37–45, 1968.

[2] ARQUÈS (Didier) et MICHELIN (Sylvain), «Régularité des objets définissant une scène et optimisation de la méthode des radiosités : application aux surfaces réglées», dans *Deuxièmes journées de l'Association Française d'Informatique Graphique*, p. 79–90, Toulouse, décembre 1994.

[3] ARVO (James), « The Irradiance Jacobian for Partially Occluded Polyhedral Sources », dans *Computer Graphics Proceedings*, Annual Conference Series, p. 343–350, (ACM SIGGRAPH '94 Proceedings), 1994.

[4] ARVO (James), TORRANCE (Kenneth), et SMITS (Brian), «A Framework for the Analysis of Error in Global Illumination Algorithms », dans *Computer Graphics Proceedings*, Annual Conference Series, p. 75–84, (ACM SIGGRAPH '94 Proceedings), 1994.

[5] BAUM (Daniel R.), RUSHMEIER (Holly E.), et WINGET (James M.), «Improving Radiosity Solutions Through the Use of Analytically Determined Form-Factors», *Computer Graphics (ACM SIGGRAPH '89 Proceedings)*, 23, n° 3, juillet 1989, p. 325–334.

[6] COHEN (Michael F.), CHEN (Shenchang Eric), WALLACE (John R.), et GREENBERG (Donald P.), «A Progressive Refinement Approach to Fast Radiosity Image Generation», *Computer Graphics (ACM SIGGRAPH '88 Proceedings)*, 22, n° 4, août 1988, p. 75–84.

[7] COHEN (Michael F.) et WALLACE (John R.), *Radiosity and Realistic Image Synthesis*, Boston, MA, Academic Press Professional, 1993.

[8] DESCARTES (René), *Discours de la Méthode pour bien conduire sa raison et chercher la vérité dans les sciences plus La Dioptrique, Les Météores et La Géométrie qui sont des essais de cette Méthode*, Leyde, Jean Maire, 1637, Republié notammment dans René Descartes, *Œuvres et Lettres*, Bibliothèque de la Pléiade.

[9] DRETTAKIS (George), Simplifying the Representation of Radiance from Multiple Emitters, dans SAKAS (Georgios), SHIRLEY (Peter), et MÜLLER (Stefan), éditeurs, *Photorealistic Rendering Techniques*, Focus on Computer Graphics, p. 259–272, Berlin, Springer, 1995 (actes du *Fifth Eurographics Workshop on Rendering*, Darmstadt, Allemagne, 13-15 juin 1994).

[10] _____, « Structured Sampling and Reconstruction of Illumination for Image Synthesis », CSRI Technical Report 293, Department of Computer Science, University of Toronto, Toronto, Ontario, janvier 1994.

[11] DRETTAKIS (George) et FIUME (Eugene), «Concrete Computation of Global Illumination Using Structured Sampling », dans *Third Eurographics Workshop on Rendering*, p. 189–202, Bristol (UK), mai 1992.

[12] _____, « Accurate and Consistent Reconstruction of Illumination Functions Using Structured Sampling», *Computer Graphics Forum (Eurographics '93)*, 13, n° 3, septembre 1993, p. 273–284.

[13] _____, «A Fast Shadow Algorithm for Area Light Sources Using Backprojection», dans *Computer Graphics Proceedings*, Annual Conference Series, p. 223–230, (ACM SIGGRAPH '94 Proceedings), 1994.

[14] _____, Structure-Directed Sampling, Reconstruction, and Data Representation for Global Illumination, dans BRUNET (Pere) et JANSEN (Frederik W.), éditeurs, *Photorealistic Rendering in Computer Graphics (Proceedings of the Second Eurographics Workshop on Rendering)*, p. 60–74, New York, Springer-Verlag, 1994.

[15] GORAL (Cindy M.), TORRANCE (Kenneth E.), GREENBERG (Donald P.), et BATTAILE (Bennett), «Modelling the Interaction of Light Between Diffuse Surfaces», *Computer Graphics (ACM SIGGRAPH '84 Proceedings)*, 18, n° 3, juillet 1984, p. 212–222.

[16] HANRAHAN (Patrick M.) et SALZMAN (David), A Rapid Hierarchical Radiosity Algorithm for Unoccluded Environments, dans BOUATOUCH (Kadi) et BOUVILLE (Christian), éditeurs, *Photorealism in Computer Rendering (Proceedings Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics, 1990)*, Eurographics Seminar Series, p. 151–171, Springer Verlag, New York, 1992.

[17] HANRAHAN (Patrick M.), SALZMAN (David), et AUPPERLE (Larry), « A Rapid Hierarchical Radiosity Algorithm », *Computer Graphics (ACM SIGGRAPH '91 Proceedings)*, 25, n° 4, juillet 1991, p. 197–206.

[18] HÉRACLITE, *Fragments*, Garnier-Flammarion, 1985.

[19] HOLZSCHUCH (Nicolas), *Vers une radiosité évoluant en temps réel*, Mémoire de DEA, DEA IMA, Université Paris XI, septembre 1992.

[20] _____, « Vers une radiosité évoluant en temps réel », dans *Journées Groplan*, Nantes, novembre 1992.

[21] _____, « Les ondelettes en synthèse d'image », dans *Colloquium* IMAG *sur les ondelettes*, Grenoble, janvier 1994, IMAG.

[22] _____, *Synthèse d'images et radiosité hiérarchique*, Mémoire de magistère, Magistère de Mathématiques Fondamentales et Appliquées et d'Informatique, École Normale Supérieure, Université Paris XI, novembre 1994.

[23] HOLZSCHUCH (Nicolas), SILLION (François), et DRETTAKIS (George), An Efficient Progressive Refinement Strategy for Hierarchical Radiosity, dans SAKAS (Georgios), SHIRLEY (Peter), et MÜLLER (Stefan), éditeurs, *Photorealistic Rendering Techniques*, Focus on Computer Graphics, p. 343–357, Berlin, Springer, 1995 (actes du *Fifth Eurographics Workshop on Rendering*, Darmstadt, Allemagne, 13-15 juin 1994).

[24] HOLZSCHUCH (Nicolas) et SILLION (François), Accurate Computation of the Radiosity Gradient for Constant and Linear Emitters, dans HANRAHAN (Patrick M.) et PURGATHOFER (Werner), éditeurs, *Rendering Techniques '95, proceedings of the Eurographics Workshop in Dublin, Ireland, June 12-14 1995*, Computer Science, p. 186–195, Wien, New-York, Springer, septembre 1995.

[25] KAJIYA (James T.), «The Rendering Equation», *Computer Graphics (ACM SIGGRAPH '86 Proceedings)*, 20, n° 4, août 1986, p. 143–150.

[26] LISCHINSKI (Daniel), SMITS (Brian), et GREENBERG (Donald P.), «Bounds and Error Estimates for Radiosity », dans *Computer Graphics Proceedings*, Annual Conference Series, p. 67–74, (ACM SIGGRAPH '94 Proceedings), 1994.

[27] LISCHINSKI (Daniel), TAMPIERI (Filippo), et GREENBERG (Donald P.), «Combining Hierarchical Radiosity and Discontinuity Meshing», dans *Computer Graphics Proceedings*, Annual Conference Series, p. 199–208, (ACM SIGGRAPH '93 Proceedings), 1993.

[28] MAX (Nelson L.) et ALLISON (Michael J.), Linear Radiosity Approximation Using Vertex-to-Vertex Form Factors, dans KIRK (David), éditeur, *Graphics Gems III*, p. 318–323, San Diego, Academic Press, 1992.

[29] MICHELIN (Sylvain), *Techniques d'optimisation et d'accélération en synthèse d'image pour la méthode des radiosités*, Thèse, Université de Franche-Comté, décembre 1995.

[30] MOON (Parry) et SPENCER (Domina Eberle), *The Scientific Basis of Illuminating Engineering*, New York, NY, McGraw-Hill, 1936.

[31] SARMANT (Jean-Pierre), *Cours de Sciences-Physiques (Classe de Mathématiques Spéciales)*, Paris, s.e., 1989.

[32] SCHRÖDER (Peter), GORTLER (Steven J.), COHEN (Michael F.), et HANRAHAN (Patrick M.), «Wavelet Projections for Radiosity », *Computer Graphics Forum*, 13, n° 2, juin 1994, p. 141–151.

[33] SCHRÖDER (Peter) et HANRAHAN (Patrick M.), «A Closed Form Expression for the Form Factor Between Two Polygons », Rapport Technique CS-TR-404-93, Department of Computer Science, Princeton University, Princeton, New Jersey, USA, janvier 1993.

[34] _____, « On the Form Factor Between Two Polygons », dans *Computer Graphics Proceedings*, Annual Conference Series, p. 163–164, (ACM SIGGRAPH '93 Proceedings), 1993.

[35] SCHWARTZ (Laurent), *Analyse II, calcul différentiel et équations différentielles*, Paris, Hermann, 1992.

[36] SILLION (François), «Clustering and Volume Scattering for Hierarchical Radiosity Calculations», dans SAKAS (Georgios), SHIRLEY (Peter), et MÜLLER (Stefan), éditeurs, *Photorealistic Rendering Techniques*, Focus on Computer Graphics, p. 105–117, Berlin, 1995 (actes du *Fifth Eurographics Workshop on Rendering*, Darmstadt, Allemagne, 13-15 juin 1994), Springer.

[37] _____, «A Unified Hierarchical Algorithm for Global Illumination with Scattering Volumes and Object Clusters», *IEEE Transactions on Visualization and Computer Graphics*, 1, n° 3, septembre 1995, p. 240–254.

[38] SILLION (François) et DRETTAKIS (George), «Feature-Based Control of Visibility Error: A Multiresolution Clustering Algorithm for Global Illumination», dans *Computer Graphics Proceedings*, Annual Conference Series, p. 145–152, (ACM SIGGRAPH '95 Proceedings), 1995.

[39] SILLION (François) et PUECH (Claude), *Radiosity and Global Illumination*, San Francisco, CA, Morgan Kaufmann, 1994.

[40] SMITS (Brian E.), ARVO (James R.), et SALESIN (David H.), «An Importance-Driven Radiosity Algorithm», *Computer Graphics (ACM SIGGRAPH '92 Proceedings)*, 26, n° 4, juillet 1992, p. 273–282.

[41] TELLER (Seth J.), «Computing the Antipenumbra of an Area Light», *Computer Graphics (ACM SIGGRAPH '92 Proceedings)*, 26, n° 4, juillet 1992, p. 139–148.

[42] TELLER (Seth J.) et HANRAHAN (Patrick M.), «Global Visibility Algorithms for Illumination Computations», dans *Computer Graphics Proceedings*, Annual Conference Series, p. 239–246, (ACM SIGGRAPH '93 Proceedings), 1993.

[43] TROUTMAN (Roy) et MAX (Nelson L.), «Radiosity Algorithms Using Higher Order Finite Element Methods», dans *Computer Graphics Proceedings*, Annual Conference Series, p. 209–212, (ACM SIGGRAPH '93 Proceedings), 1993.

[44] VEDEL (Christophe), «Improved Storage and Reconstruction of Light Intensities on Surfaces», dans *Third Eurographics Workshop on Rendering*, p. 113–121, Bristol, UK, mai 1992.

[45] WARD (Gregory J.) et HECKBERT (Paul), «Irradiance Gradients», dans *Third Eurographics Workshop on Rendering*, p. 85–98, Bristol, UK, mai 1992.

[46] WARUSFEL (André), *Cours de Mathématiques (Classe de Mathématiques Spéciales)*, Paris, s.e., 1989.

[47] WHITTED (Turner), «An Improved Illumination Model for Shaded Display», *CACM*, 23, n° 6, juin 1980, p. 343–349.

[48] ZATZ (Harold R.), «Galerkin Radiosity: A Higher Order Solution Method for Global Illumination», dans *Computer Graphics Proceedings*, Annual Conference Series, p. 213–220, (ACM SIGGRAPH '93 Proceedings), 1993.