Reprezentace funkce a sférické harmonické

Aproximace funkce

V této kapitole se budeme zabývat tím, jak aproximovat funkci konečnou množinou reálných čísel a jak tuto dovednost využít při výpočtu globálního osvětleni scény z mapy prostředí v reálném čase. Při renderingu obrázků osvětlených mapou prostředí (tzv. Environment Map - zkr. EM) často narážíme na časovou náročnost výpočtu osvětleni na povrchu scény, zejména u difuzních ploch, kde je třeba integrovat světlo dopadající z EM pres celou hemisféru. Jedním ze způsobů, jak snížit počet operací spojených s osvětlením mapou prostředí je právě zmenšení počtu vstupních dat (zde máme na mysli právě mapu prostředí) při co největším zachování informace, tedy vhodnou aproximaci. Aproximovanou funkci (EM) pak vyjádříme ve formě několika (relativně) málo reálných čísel. V takovéto reprezentaci budeme přímo provádět výpočet osvětlení na povrchu scény, tudíž dosáhneme značného omezení (dle aproximace) počtu potřebných operací. Např. pokud se mi povede mapu prostředí o velikosti v řádech megapixelů reprezentovat vektorem o velikosti 3 reálných čísel při zachování kvality osvětlení, dostanu řádovou úsporu při práci v takové reprezentaci.

Platí, že každou L2 integrovatelnou funkci lze zapsat jako lineární kombinaci nějakých bázových funkcí. Tedy :

$$G(x) = \sum_i c_i B_i(x)$$

Kde c_i jsou koeficienty a B_i jsou dané bázove funkce.

Typickým příkladem takového zápisu funkce je Taylorův rozvoj. Pokud omezíme počet sčítanců této sumy na počet menší (budeme tedy sčítat pouze prvních n sčítanců) bude se jednat o aproximaci původní funkce $G(x)$. Tato vlastnost neplatí pouze u Taylorovy řady, ale je obecnou vlastností takovéto reprezentace funkce. Mnoho systémů (Taylor, Fourier, wavlet, ...) je navrženo tak, aby několik málo prvních koeficientů neslo co nejvíce informací (tj. aby aproximace byla co nejpodobnější původní funkci)[1] o dané funkci. Toho právě často využívají metody komprese, aproximace nebo jen řídké reprezentace. V takových případech lze tedy reprezentovat danou funkci pouze jedním vektorem o délce n, jehož prvky jsou jednotlivé koeficienty. To samozřejmě platí pouze za podmínky, že máme pevně zvolené bázové funkce. Pro naše účely zde budeme uvažovat pouze koeficienty reálné.

[1] Zde záleží na definici podobnosti (př. vjemová u obrazových dat, nebo vzdálenost v nějaké metrice pro funkce.)

Pro aproximaci libovolné funkce pomoci koeficientů bázových funkcí je nejprve třeba tyto koeficienty najít. Koeficienty budeme značit c_i ; aproximovanou funkci $G(x)$. Snadno nahlédneme, že pro zvolené bázové funkce lze hledané koeficienty najít pomocí minimalizace chyby (mezi vyjádřením funkce pomocí lineární kombinace bázových funkcí a původní funkcí $G(x)$) v L_2 norme, tedy :

$$E_{L_2} = \int_I [G(x) - \sum c_i B_i(x)]^2 dx$$

Kde E_{L_2} je celková chyba, kterou jsme se aproximací dopustili měřená v norme L_2, $G(x)$ je originální funkce a $\sum c_i B_i(x)$ jeji aproximace.

Euler Lagrangeova rovnice nám říká, že podmínkou pro nalezeni minima teto funkce je nulovost derivace dle c_i. Vztah lze po zderivování zapsat maticově jako :

$$\boldsymbol{Bc} = \boldsymbol{G}$$

Kde : $\boldsymbol{B} = \begin{bmatrix} < B_1|B_1 > & \cdots & < B_1|B_n > \\ \vdots & \ddots & \vdots \\ < B_n|B_1 > & \cdots & < B_n|B_n > \end{bmatrix}$;

$\boldsymbol{c}$ je vektor koeficientů;

$\boldsymbol{G} = [< G|B_1 > \ \dots \ < G|B_n >]^T$ a

Výraz $< F|H >$ značí skalární součin daný předpisem $< F|H > = \int_I F(x)H(x)dx$.

Ze vztahů uvedených výše již snadno nahlédneme, že matice $\boldsymbol{B}$ závisí pouze na zvolených bázových funkcích B_i a vůbec nezávisí na $G(x)$. Lze tedy matici $\boldsymbol{B}$ předpočítat dopředu pro danou bázi.

Hledané koeficienty c_i tedy snadno nalezneme pomocí: $\boldsymbol{c} = \boldsymbol{B}^{-1}\boldsymbol{G}$. Zde by se nam ale velmi hodilo, aby námi zvolené báze byla ortonormální. Ortonormální báze mají totiž tu hezkou vlastnost, že její jednotlivé složky jsou na sebe navzájem kolmé. To mimo jiné znamená, že skalární součin libovolných dvou navzájem různých bází je vždy roven nule. Tedy:

$$< B_i|B_j > = \{ {}^{1\ |\ i=j}_{0\ |\ jinak}$$

Výše uvedený vztah hledání koeficientů lze tak v případě ortonormální báze zjednodušit na: $\boldsymbol{c} = \boldsymbol{G}$, čímž se nám hledání velmi zjednoduší. Pro usnadnění práce v takovémto systému ještě uvedeme následující důležitý vztah:

$$\int f(x)g(x)dx = \int [\sum f_i B_i(x)].\left[\sum g_j B_j(x)\right]dx$$

$$= \int \sum\sum f_i g_j B_i B_j dx$$

$$= \sum\sum f_i g_j \int B_i B_j dx$$

$$= \sum f_i g_i$$

Pozn. Integrál $\int B_i B_j dx$ se v posledním kroku promění v Kroneckerovu delta funkci.

Tento vztah vlastně znamená, že pro ortonormální bazický systém je integrál součinu dvou funkcí roven skalárnímu součinu koeficientů jejich bazických funkcí. Tato vlastnost se nám bude velmi hodit, protože rovnice odrazu na povrchu je právě v takovémto tvaru.

Výše popsané vlastnosti jsou natolik důležité, že je nadále zbytečné uvažovat jiné než ortonormální bázové systémy.

Sférické harmonické funkce

Příkladem ortonormálního bázového systému jsou sférické harmonické funkce (SH). Sférické harmonické funkce jsou ortogonální řešení úhlové části Laplaceovy rovnice vyjádřená ve sférickém souřadném systému. Jednotlivé funkce se obvykle značí

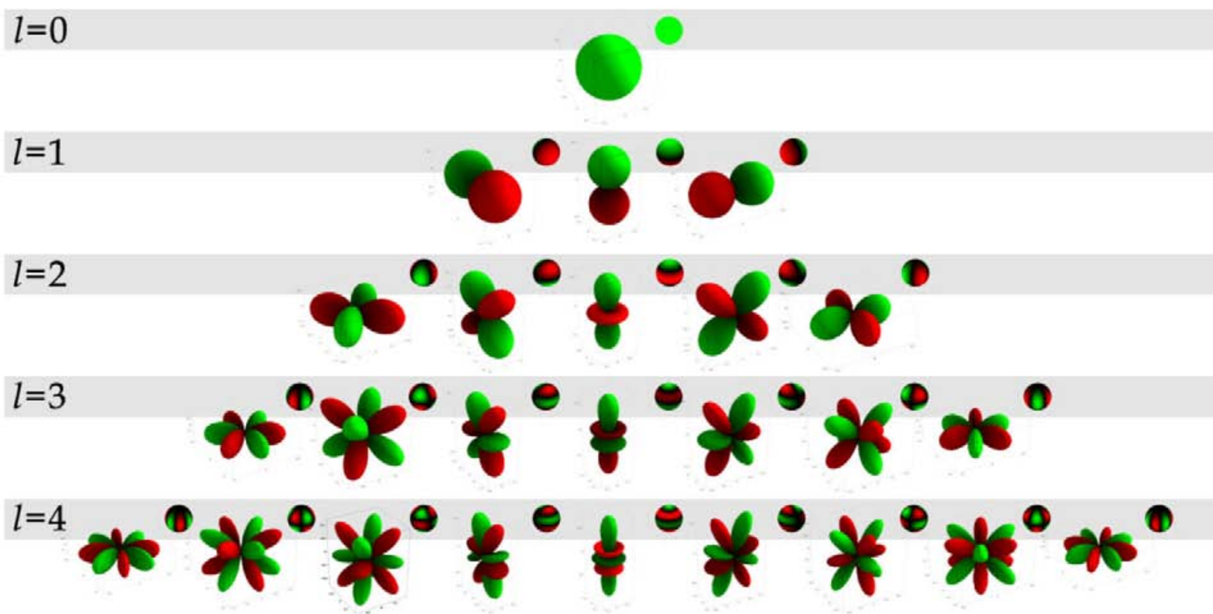
$Y_{lm}(\theta, \varphi)$, případně $Y_i(\theta, \varphi)$, kde $i = l(l+1) + m$,

t.ž. $l \geq 0; m \in <-l, +l>$

Parametry l a m jsou celočíselné proměnné značící stupeň resp. řád řešeného Lagrageova polynomu. Jejich skutečný význam však není pro naši potřebu podstatný.

Parametry (θ, φ) značí úhel ve sférickém souřadném systému.

V podstatě se ale jedná o soubor bazických funkcí definovaných na povrchu jednotkové koule. Obrázek níže poslouží pro obecnou představu, jak tyto báze vypadají. Vzorce SH funkcí jsou dostupné v literatuře a my je zde uvádět nebudeme.



Pomocí těchto harmonik budeme aproximovat funkci $G(\theta, \varphi)$ způsobem zmíněným v předchozí kapitole. Tedy

$$G(\theta, \varphi) = \sum_{l \geq 0} \sum_{m \in <-l, +l>} c_{l,m} Y_{l,m}(\theta, \varphi)$$

Pro účely výpočtu globálního osvětlení budeme sférické harmonické používat pro aproximaci osvětlení z mapy prostředí. V naší úloze se nám bude jednat o urychlení výpočtu irradiance lambertovského povrchu osvětleného mapou prostředí. Tato je při zanedbání stínů pouze funkcí normály povrchu a ničeho jiného. Jedná se tedy o funkci definovanou na povrchu jednotkové koule a jako takovou ji můžeme reprezentovat právě pomocí koeficientů SH.

Realtime výpočet osvětlení z mapy prostředí.

Sférické harmonické

Výpočet globálního osvětlení z mapy prostředí je snadný pro ideálně odrazivé povrchy, ale pro matné až lambertovské plochy je třeba integrovat z mapy prostředí pres celou hemisféru určenou normálou na povrchu. Jelikož je tento výpočet paměťově i časově náročný, budeme se jej snažit zjednodušit pomocí aproximace výpočtu irradiance povrchu několika mála reálnými koeficienty. Při zanedbání stínů platí, jak bylo zmíněno v minulé kapitole, že výpočet irradiance na povrchu osvětleném mapou prostředí je pouze funkcí normály a lze jej tedy vyjádřit pomocí koeficientů sférických harmonických funkcí.

Dále vyjdeme z toho, že osvětlení lambertovských povrchů z mapy prostředí je ekvivalentní odrazu a rozmazaní mapy prostředí a lze jej tedy analyticky zapsat pomocí koeficientů SH jako:

$$E_{lm} = A_l L_{lm}$$

Kde E_{lm} jsou koeficienty rozmazané (přefiltrované) mapa prostředí; A_l je low-pass filtr a L_{lm} je mapa prostředí vyjádřena pomocí koeficientů SH:

$$L_{lm} = \int EM(\omega) Y_{lm}(\omega) d\omega = \sum EM_{ij} Y_{lm}(\omega_{ij}) \Delta ij \ ,$$

kde EM znaci mapu prostredi, Y_{lm} jsou bazove funkce SH ;

Nutno poznamenat, že pro l>3 se A_l velmi blíží nule, a platí tedy, že aproximace mapy prostrdi prvnimi tremi koefifienty je pro výpočet irradiance lambertovských povrchů zcela dostačující.

Realtime rendering

Při renderingu je třeba vyhodnocovat dotazy pro radianci přicházející z daného směru. Jelikož používáme pro osvětlení reprezentaci pomoci sférických harmonických funkcí, vyhodnocení dotazu lze zapsat následovně:

$$E(n) = \sum E_{lm} Y_{lm}(\omega)$$

Kde E_{lm} jsou právě koeficienty SH reprezentující irradianci z mapy prostředí na povrchu scény a Y_{lm} jsou príslušné bázové funkce. V našem případě, při použití low-pass filtru popsaného výše, tj. použití koeficientů pouze do řádu 3 lze vztah přepsat následujícím způsobem:

$$E(\omega) = \sum E_{lm} Y_{lm}(\omega) = \sum_{l=0}^{2} \sum_{m=-l}^{l} E_{lm} Y_{lm}(\omega)$$

Sférické harmonické funkce do řádu 3 jsou maximálně kvadratickými polynomy kartézských souřadnic směru normály. Protože platí, že všechny nejvýše kvadratické polynomy lze popsat pomocí kvadratické formy v homogenních souřadnicích lze uvedený vztah obecně přepsat ve formě:

$$E(n) = n^t M n$$

Kde matici M lze predpoocitat dopredu.

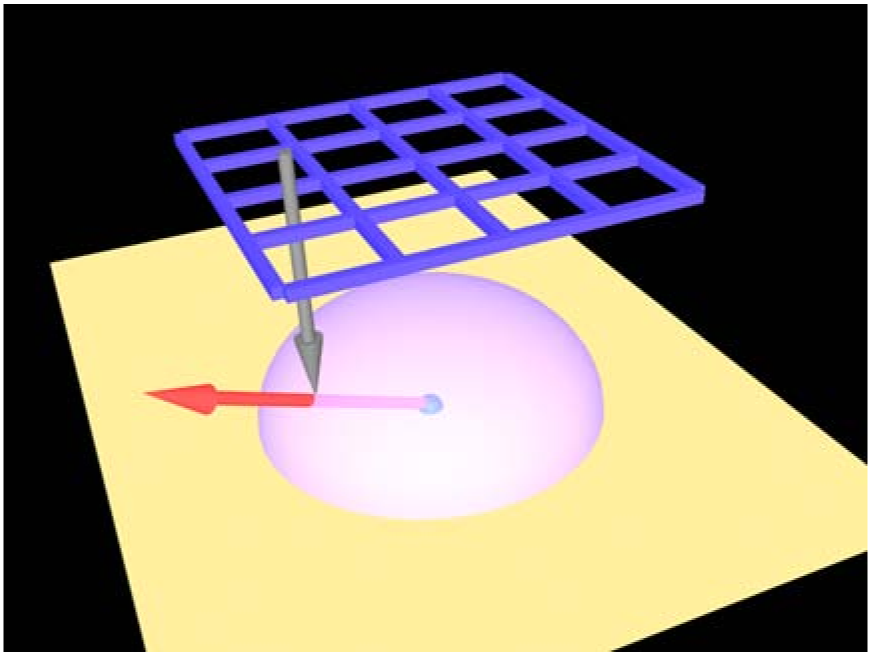
Počítání s obecnou BRDF

Až doposud jsme se zabývali výpočtem osvětlení na lambertovském povrchu, kde platí, že odchozí radiance není závislá na směru pohledu.

Pokud bychom chtěli použít aproximace pomoci sférických harmonických funkcí na povrchy s obecnou BRDF, pak využijeme toho, že pro zafixovaný odchozí směr je i BRDF funkcí definovanou na sféře. Lze ji tedy pro každý směr opět reprezentovat pomoci SH způsobem popsaným v předchozí kapitole. Na rozdíl od lambertovských povrchů nám zde zcela jistě nepostačí reprezentovat EM pomocí pouhých tří koeficientů. Jejich počet bude záviset na lesklosti povrchu (intuitivně lze říci, že odraz od povrchu bude méně rozmazán –> větší detail –> více koeficientů). Výpočet odrazu lze pro urychlení výpočtu zapsat jako skalární součin koeficientu SH jak jsme si ukázali v první části lekce.

$$L_o(\omega_o) = \int_{\Omega} L_i(\omega_i) BRDF(\omega_i, \omega_o) cos(\theta_i) d\omega_i = \int EM.BRDF$$

Pro využití SH je tedy pro danou BRDF třeba předpočítat koeficienty SH pro všechny odchozí směry paprsku (s využitím rotační invariance u izotop. povrchu) a tuto informaci uchovat například v textuře. Pro tento účel použijeme nejlépe parabolické mapování (viz obr.).



Ukázka parabolického mapování textury na sféru.

Problémem tohoto přístupu ale je, že mapa prostředí je typicky definována v globálním souřadném systému, na rozdíl od BRDF, která je naopak definována v souřadném systému lokálním. Řešením je aplikace rotace při výpočtu odrazu tak aby byla EM převedena do souřadného systému povrchu, nebo naopak. Zde lze využít toho, ze SH jsou uzavřené na rotaci a výpočet této rotace lze tedy jednoduše reprezentovat maticovým součinem.