

MOŽNOSTI A VYLEPŠENÍ SIMULAČNÍ METODY LHS

LIMITATIONS OF AND IMPROVEMENTS TO LHS

Miroslav Vořechovský¹

Abstract

A new technique is proposed to improve the performance of Latin hypercube sampling. To reduce error in correlations between variables (correlated or uncorrelated), it is proposed to perform Simulated Annealing method on the realizations for all variables, instead of using matrix manipulation, as is the current custom. Limitations in hypercube sampling will be discussed. To improve the statistics for each variable, it is proposed that realizations for each variable be acquired by finding the probabilistic means of equiprobable disjoint intervals in the variable's domain, instead of using the cumulative distribution function directly, as is currently done.

1 Úvod

Simulace typu Monte Carlo (MC) je mocný nástroj, velmi často používaný k analýze náhodných jevů pomocí počítačů. U náhodných problémů je požadována statistická či pravděpodobnostní informace o náhodném výstupu (odezvě) závisující na náhodných vstupních veličinách, polích a procesech. Simulace MC je robustní, snadno použitelná a obecně rychlejší než plně pravděpodobnostní přístupy a proto je často používána pro řešení náhodných problémů a pro ověřování jiných metod analýzy. Metoda Latin Hypercube Sampling (LHS) představuje jeden z nejúčinnějších postupů k odhadu statistických parametrů odezvy (konstrukce) a pro tento účel byla také původně vyvinuta [3]. Výhodou je možnost výrazného snížení počtu simulací oproti jednoduché metodě MC při zachování přesnosti odhadů statistik odezvy.

Nabízená úprava metody LHS zachovává funkci rozdělení všech simulovaných veličin a dodržuje požadované korelace mezi veličinami. LHS sestavuje vysoce provázanou sdruženou hustotu pravděpodobnosti vektoru náhodných veličin, která umožňuje vysokou přesnost statistických parametrů odezvy při použití malého počtu simulací. Lze rozlišit dvě fáze simulace LHS. Nejdříve jsou vybrány vzorky reprezentující funkce hustoty pravděpodobnosti veličin. Poté se pracuje s pořadími vzorků tak, aby se dosáhlo požadované statistické závislosti mezi veličinami. Vzhledem k tomu, že statistická závislost je vzorkům vnucena pouze záměny pořadí a nikoliv změnou jejich hodnot, rozdělení pravděpodobnosti veličin zůstane nezměněno.

2 Metoda Latin Hypercube Sampling

V následujícím uvedeme hlavní myšlenku metody. Odhady statistických parametrů rezervy spolehlivosti jsou získány z určitého počtu simulovaných realizací funkce mezního stavu podobně jako v jednoduché metodě Monte Carlo. Definiční obor funkce hustoty $f(X_i)$ každé základní náhodné veličiny X_i je totiž rozdělen na N_{Sim} disjunktních intervalů. Volí se jednoduše intervaly o stejné pravděpodobnosti $1/N_{Sim}$. Jedná se o stratifikační metodu, kde vrstva

¹ Miroslav Vořechovský, Ing., Vysoké učení technické v Brně, Fakulta stavební, Ústav stavební mechaniky, Veveří 95, 662 37 Brno, tel.: 41 14 71 31, E-mail: vozechovsky.m@fce.vutbr.cz

oboru hodnot distribuční funkce náhodné veličiny je nahrazena jedinou hodnotou. Je zajištěno, že celý rozsah každé základní náhodné veličiny je realizován při simulaci rovnoměrně vzhledem k distribuční funkci. Současně dosáhneme toho, že žádná reálná hodnota není předem vyloučena, každá vrstva distribuční funkce se při simulaci použije právě jednou. Tyto skutečnosti pak vedou k velmi dobrým odhadům při poměrně nízkém počtu simulací [3].

3 Výběr vzorků

První fáze metody LHS spočívá ve výběru vzorků reprezentujících intervaly rozdělení pravděpodobnosti náhodné veličiny. Stávající praxe je vybírat vzorky přímo inverzní transformací pravděpodobnostní funkce, a to v polovině vrstvy rovnoměrně rozděleného oboru hodnot kumulativní distribuční funkce. Kritizovat lze redukci výběru náhodných veličin z dané vrstvy na střed intervalu. Námítka se týká zvláště krajních vrstev, které nejvíce ovlivňují šikmost a špičatost rozdělení výstupních veličin. Uvedený postup u symetrických rozdělení poskytuje přesnou střední hodnotu získaného souboru vzorků (odhadovanou aritmetickým průměrem) a obecně je střední hodnota blízká požadované. Ovšem rozptyl vzorků je zpravidla značně rozdílný od požadovaného. Lépe je provádět výběr vzorků jako střední hodnotu každého intervalu [7], tedy s pomocí přídavné integrace. Vzorky pak lépe reprezentují funkce hustoty pravděpodobnosti veličiny. Střední hodnota souboru je získána přesně (aditivní vlastnost integrálu) a rozptyl dat je mnohem bližší požadovanému. Pro některá rozdělení pravděpodobnosti (včetně normálního) je integrál v [7] snadno řešitelný analyticky. V případě, že nalezení primitivní funkce není triviální, je nutno použít dodatečný numerický výpočet integrálu, ovšem takový nárůst výpočtové náročnosti se jistě vyplatí. Vzorky vybrané oběma popsanými způsoby jsou téměř identické až na „konce“ oboru hodnot veličiny. Proto lze při simulaci veličiny s konečným oborem hodnot použít jednodušší postup (v polovinách vrstev). Pokud je obor hodnot rozdělení veličiny nekonečný či semi-nekonečný a numerická integrace přináší obtíže, stačí použít složitější výběr ze [7] pouze u několika prvních a poslední vzorků. I v takovém případě budou spočtené statistiky souboru vzorků velmi blízko požadovaným.

4 Zavedení statistické závislosti, odhad korelačních koeficientů

Poté co jsou vzorky všech veličin vygenerovány, je nutno pracovat se statistickou korelací mezi veličinami (nulovou či jinou). Dopusud bylo předpokládáno, že vygenerované vektory matice \mathbf{X} jsou nezávislé. Tomu by ovšem měla odpovídat i nezávislost statistických souborů tvořících jednotlivé realizace veličiny. Existuje možnost, že mezi sloupci vznikne nezanedbatelná statistická závislost, která by mohla výrazně ovlivnit získané výsledky. O tom, zda matice \mathbf{X} skutečně splňuje předpoklady statistické nezávislosti se můžeme přesvědčit např. pomocí Spearmanova koeficientu pořadové korelace nebo s pomocí klasického lineárního korelačního koeficientu. Je vhodné korelační koeficienty mezi veličinami měnit pouze změnami pořadí vzorků u jednotlivých veličin a neměnit již hodnoty vzorků. Nalezení výkonné metody záměn pořadí vzorků u každé veličiny umožní zavedení požadované statistické závislosti mezi nimi. Zde není nutno diskutovat okolnost, že nalezení sdružené rozdělovací funkce hustoty pravděpodobnosti pro všechny jednotlivé složky náhodného vektoru respektující požadované korelační koeficienty je v obecném případě nemožné. Pokud by to bylo možné (např. mnohorozměrné normální rozdělení), výběr reprezentantů by se realizoval přímo z takové sdružené funkce hustoty pravděpodobnosti.

4.1 ULHS a iterační metody

Uvažme případ, že všechny veličiny mají být generovány jako nekorelované. Doposud bylo možno vzorky generovat v náhodném pořadí uvažovat statistické závislosti jako dostatečně malé. Ovšem takto by pravděpodobně byla korelace mezi některými veličinami významná [1]. Florian [1] pro redukci náhodně zavedené korelace navrhuje metodu Updated Latin Hypercube Sampling (ULHS). Jeho postup je založen na technice, kterou lze nalézt v publikaci [3] pro zavedení požadované korelace mezi hypoteticky nekorelovanými veličinami. Myšlenka metody ULHS je podrobně rozebrána např. v autorově publikaci [7]. Uvedme pouze fakt, že v metodě ULHS jsou pro třídící postup používána pouze pořadová čísla, nikoliv samotné vzorky. Korelace mezi sloupci v matici pořadových čísel je pak s výhodou definována Spearmanovým korelačním koeficientem [1], [7]. Bylo prokázáno [3], že korelace mezi jakýmkoliv dvěma veličinami systému se nezvyšuje a může se snižovat. Poznamenejme, že tato technika může být provozována iterativně a teoreticky dovolí značně snížit korelaci. Ovšem v praxi autoři zjišťují, že ULHS má tendenci konvergovat k pořadím, kde stále mezi některými veličinami zůstává významná chyba v korelaci. Proto je pro odstraňování nechtěné korelace vhodné použít jinou techniku.

Je-li požadováno simulovat korelované veličiny, pak má zmíněný postup jisté potíže. Nejdříve je nutno použít ULHS pro minimalizaci korelace. Pak je matice \mathbf{K} s požadovanými korelačními koeficienty podrobena Choleského dekompozici [1], [7] a pak je získána nová matice pořadí. Tato procedura pro zavádění korelace může být použita pouze jednou, není zde možnost zlepšování iteracemi. Nová matice pořadí nedovoluje veličinám dobrou shodu s požadovanými korelačními koeficienty a pro zlepšení korelace nelze nic udělat. Je potřeba najít jinou metodu pro simulaci jak nekorelovaných, tak korelovaných veličin.

4.2 Spektrální rozklad korelační matice

V rámci hledání metody pro zavedení požadované korelace byla studována možnost provedení spektrálního rozkladu požadované korelační matice

$$\mathbf{K} = \mathbf{\Phi} \cdot \mathbf{\Lambda} \cdot \mathbf{\Phi}^T \quad (1)$$

kde $\mathbf{\Phi}$ je ortogonální transformační matice (matice vlastních vektorů \mathbf{K}) a diagonální matice $\mathbf{\Lambda}$ obsahuje (kladná, reálná) vlastní čísla \mathbf{K} reprezentující korelační matici v nekorelovaném prostoru veličin. Matice simulací by se pak mohla získat následovně:

$$\mathbf{X} = \mathbf{\Phi} \cdot \mathbf{K} \quad (2)$$

Uvedený postup by modifikoval hodnoty vzorků a navíc získané korelační koeficienty matice \mathbf{X} se významně liší od požadovaných. Použije-li se získaná matice \mathbf{X} jen k přetřídění původních vzorků ve stejném pořadí, korelační koeficienty se příliš nezlepší.

4.3 Provedení všech možností kombinací pořadí vzorků

Nejpřesnější možností zavedení požadované korelace je vyzkoušení všech kombinací pořadí vzorků pro jednotlivé veličiny. Ponecháme-li jeden sloupec vzorků nepermutovaný, je nutno vyzkoušet značné množství uskupení $N = (N_{sim} - 1)^{N_v - 1}$. Lze snadno výpočtem ověřit, že pro reálné úlohy je dnes i v blízké budoucnosti uvedený postup nerealizovatelný. Slouží ovšem jako test jiných metod pro určení, zda bylo nalezeno nejlepší řešení (vzájemné pořadí) u malých úloh.

4.4 Optimalizační metoda simulovaného žíhání

Pro simulaci korelovaných i nekorelovaných metod byla v rámci výzkumu vypracována metoda s použitím optimalizačního postupu simulovaného žíhání. Předkládaná metoda může pracovat jak se Spearmanovým koeficientem, tak s klasickým lineárním či jiným koeficientem korelace. Pro každou dvojici veličin je vypočten aktuální korelační koeficient S_{ij} . Poté lze zvolit vhodnou míru či normu E pro hodnocení kvality celkových statistických závislostí v matici \mathbf{X} . Jako vhodný ukazatel přesnosti se jeví např. maximální rozdíl v korelační matici požadované \mathbf{K} a aktuální \mathbf{S} :

$$E = \text{abs} \left(\max_{1 \leq i, j \leq N_V} (S_{i,j} - K_{i,j}) \right), \quad (3)$$

nebo norma, která celkově zohlední deviace všech koeficientů s druhou mocninou:

$$E = \sum_{i=1}^{N_V} \sum_{j=1}^{N_V} (S_{i,j} - K_{i,j})^2 \quad (4)$$

Zde lze samozřejmě doporučit výpočet rozdílů pouze pro horní nad-diagonální trojúhelníky symetrických korelačních matic.

Prvky normované matice \mathbf{U} jsou nyní přetřídřovány podle algoritmu simulovaného žíhání, a to vždy tak, že se provede pouze jedna záměna dvojice vzorků u jedné veličiny najednou. To umožní také značné úspory při přepočtu korelační matice. Stačí znát pozice vyměněných vzorků a korelační matici upravovat jen pro dotčené korelační koeficienty (v řádku a sloupci týkající se dotčené veličiny). Na konci simulačního procesu mohou pokračovat výpočty odezvy s jednotlivými simulacemi (řádky) matice \mathbf{X} . Zavádění požadované korelace je zde pojato jako optimalizační proces, ve kterém se minimalizuje jedna z norem (3), (4). Jedná se vlastně o hledání absolutního minima diskrétní funkce mnoha proměnných.

Při výzkumu metod byla zkušebně použita jednoduchá stochastická metoda, která náhodně vymění pořadí dvojice vzorků náhodně vybrané veličiny a v případě, že došlo ke snížení chyby, byly matice \mathbf{U} přijata jako nová generace, tedy jako výchozí matice. Takový postup ovšem v drtivé většině případů skončí ve stavu, kdy již jakákoliv výměna dvojice vzorků nepřináší snížení normy E a tudíž nemůže být realizována a přijata. Lze říci, že metoda nachází určitá lokální minima. Poznamenejme, že za globální minimum lze považovat v tomto smyslu výsledek řešení podle odst. 4.3, realizovatelný pouze pro velmi malé systémy. Je tedy nutné použít metodu, která pomocí náhodných operátorů a navíc s určitou pravděpodobností neuvízne v lokálním minimu. Osvědčila se aplikace metody simulovaného žíhání [6]. Princip metody simulovaného žíhání je založen na porušení podmínky, že nová generace vzorků může být přijata pouze v případě, že došlo ke zmenšení normy (3), (4) nebo podobné. V každé iteraci dojde k vytvoření generace potomků výměnou dvojice vzorků, která se přijme či nepřijme. Simulované žíhání se aplikuje ve druhém případě. Jestliže nedojde ke zmenšení funkční hodnoty optimalizované funkce (3) či (4), je nový vektor s určitou pravděpodobností přijat. Pravděpodobnost přijetí se řídí podle Boltzmannova rozdělení pravděpodobnosti:

$$P_r(E) \approx e^{\left(\frac{-\Delta E}{k_b \cdot T} \right)} \quad (5)$$

kde ΔE je rozdíl norem E před záměnou a po záměně. Toto rozdělení vyjadřuje představu, že systém v teplotní rovnováze o teplotě T má svou energii pravděpodobnostně rozloženu mezi všechny různé možné energetické stavy E . Dokonce i při nízkých teplotách existuje určitá pravděpodobnost (velmi malá), aby byl systém lokálně ve vyšším energetickém stavu.

Proto existuje odpovídající možnost pro systém přemístit se z lokálního energetického minima ve prospěch nalezení lepšího, „více“ globálního minima. Hodnota k_b (Boltzmannova konstanta) je veličina, která uvádí do vztahu teplotu a energii systému. Metoda principiálně představuje analogii s tuhnutím krystalu. Principu simulovaného žíhání lze názorně porozumět na základě energetického grafu v [7]. Je zde znázorněno lokální minimum, a globální minimum. Představme si, že zde máme umístěnu kuličku, která může zaujmout stabilní polohu v jednom z těchto minim. Jestliže určitý algoritmus vyhledá lokální minimum, v případě deterministických metod v něm kulička většinou skončí. U simulovaného žíhání existuje určitá pravděpodobnost (5), že kulička přeskóčí do minima globálního. Systém však musí být přiměřeně excitován (zahřát), aby energie, kterou kulička potřebuje k vyskočení z lokálního minima, byla dostatečná. Analogicky si můžeme toto zahřátí představit jako zatřesení. Pokud bude zatřesení malé, pak kulička zůstane v lokálním minimu. Chceme-li dostat kuličku ven, musíme působit větší intenzitou. Pokud budeme aplikovat třes vysoké intenzity, pak bude kulička přeskakovat z minima lokálního do globálního a naopak. Je tedy zřejmé, že neoptimálnější postup je začít s vysokou intenzitou (teplotou) a postupně ji snižovat až téměř k nule a přitom si v průběhu „zapamatovat“ nejnižší polohu kuličky.

V našem případě lze hodnotu k_b nastavit jako jednotkovou (vypustit ze vztahu), protože se nabízí normu E a teplotu T uvažovat ve stejném rozměru. Zpravidla se doporučuje počáteční teplotu nastavit heuristicky tak, aby po provedení několika prvních iterací byl poměr úspěšných iterací ku všem iteracím přibližně 0,5. Tak je potřeba se zachovat v případě, kdy je metoda simulovaného žíhání použita pro hledání extrémů funkce, jejíž extrémní hodnoty nejsou vůbec známy a o vyšetřované funkci není k dispozici dostatečná tvarová představa. Ovšem problém přibližování se k dané korelační matici skutečnou korelační maticí je značně ohraničen. Přípustné členy korelační matice leží vždy v intervalu $\langle -1; 1 \rangle$. Proto lze spočítat maximum normy (3) či (4) jako jistou mocninu sumy rozdílů příslušných pozic v požadované a hypoteticky nejodlišnější korelační matici. Takto lze také inteligentně odhadovat maximální (počáteční) teplotu systému. Vidíme již jasný vztah mezi normou E a teplotou T . Teplota systému je dále monotónně snižována redukčním faktorem. Teplotní redukční faktor T_f se obecně doporučuje nastavit na hodnotu 0,95. Lze ovšem používat varianty se sofistikovanějším systémem ochlazování (cooling schedule), např. vhodné adaptivní algoritmy. Snižování počáteční teploty T a každá další úprava této veličiny teplotním redukčním faktorem T_f se provádí heuristicky např. po konstantním počtu iterací:

$$T_{i+1} = T_i \cdot T_f \quad (6)$$

Jako koncovou teplotu lze po zkušenostech doporučit velmi nízkou hodnotu, řádově okolo 10^{-8} . Výsledky testování ukazují na fakt, že při těchto nízkých teplotách dochází v procesu ladění stále k významným zlepšením ve smyslu normy (3) či (4). Vyšší počet obrátek na jedné teplotní úrovni obecně zlepšuje dosažené výsledky. Je zde ovšem nevýhoda pochopitelného nárůstu výpočtového času.

Typický průběh snižování chyby, teploty a excitace systému v průběhu žíhání je na obr. 2 v [7]. Klesající pravděpodobnost (5) se projevuje zúžením pásu okolo průměrné normy (chyby). Simulované žíhání zabraňuje v konečné fázi zpřesnění hodnot funkčního minima. Proto je výhodné po určitém počtu iterací vliv simulovaného žíhání potlačit. Po jeho ukončení přijme ta generace, při které byla funkční hodnota minimalizované funkce v celém dosavadním průběhu optimalizace nejmenší. Případný další průběh řešení představuje již pouze jakousi formu doladování a zpřesňování hodnoty minima metodou prostého náhodného výběru bez možnosti přijetí generace s větší chybou E . V [7] jsou osvětleny aspekty zavádění požadované korelace u korelačních matic, které nejsou pozitivně definitní. Je také navržena technika „vážené“ formy metody, kde se některé korelační koeficienty plní přesněji nežli jiné. Zde se také

objevuje faktor počtu simulací N_{Sim} , čím větší N_{Sim} , tím je větší operativní prostor metody a pro velké počty simulací jsou dosahovány vynikající shody s požadovanou korelační maticí.

5 Závěry

Je předložena originální technika pro zvýšení výkonnosti metody Latin hypercube sampling. Předložena je metodika snížení rozdílů mezi požadovanou korelační maticí a korelační maticí spočtenou nad realizacemi vzorků metody LHS. Je doporučena optimalizační metoda simulovaného žihání. Takový postup společně s přesnějším výběrem vzorků z libovolných rozdělovacích funkcí náhodných veličin vede k výraznému zkvalitnění metody. Výhodou metody je, že simulační proces lze kdykoliv ukončit, např. při dosažení uspokojivé míry korelace (normy). Lze použít váženou metodu zavádění korelace, kdy uživatel klade důraz na co nejpresnější splnění vybraných korelačních koeficientů, jinak se odchylky realizují rovnoměrně. Jedná se o robustní, stabilní, velmi výkonnou a rychlou metodu. Metoda LHS je obecně použitelná v libovolném oboru (sociologie, lékařství, elektroinženýrství, atd.), kde je nutno sledovat charakteristiky náhodného výstupu v závislosti na popsáném náhodném vstupu. Návazně lze provádět také citlivostní analýzy.

V současné době je autorem metoda použita ke generaci náhodných procesů metodou ortogonální transformace [5]. Lze ověřit, že se tak výrazně zlepšily statistiky generovaných polí včetně autokorelační funkce. Autor v poslední době také úspěšně použil metody zprostředkovaně ke generování cross-korelovaných náhodných polí, tedy u generování více korelovaných procesů.

PODĚKOVÁNÍ: Autor děkuje následujícím vědecko-výzkumným záměrům a grantům: CEZ: J22/98:261100007, GAČR 103/00/0603 a v neposlední řadě také školiteli Doc. Ing. Drahomíru Novákovi, DrSc. za účinné vedení při řešení úlohy.

Literatura

1. FLORIAN, A. An efficient sampling scheme: updated latin hypercube sampling. Probabilistic Engineering Mechanics, 1992, 7, 123-130.
2. HUNTINGTON, D.E. & LYRINTZIS, C.S. 1998. Improvements to and limitations of Latin hypercube sampling. Probabilistic Engineering Mechanics, 13 (4): 245-253.
3. IMAN, R. L. & CONOVER, W. J. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. Communications in Statistics: Theory and Methods, 1980, A9, 1749-1842.
4. IMAN, R.C. & CONOVER, W J. 1982. A Distribution Free Approach to Inducing Rank Correlation Among Input Variables. Communication Statistic, B11: 311-334.
5. NOVÁK, D., LAWANWISUT, W. & BUCHER C. 2000. Simulation of random field based on orthogonal transformation of covariance matrix and Latin Hypercube Sampling. Proc. of Int Conf. On Monte Carlo Simulation Method MC2000, Monaco, Monte Carlo.
6. OTTEN, R. H. J. M. & GINNEKEN, L. P. P. P. 1989. The Annealing Algorithm. Kluwer Academic Publishers, USA.
7. VOŘECHOVSKÝ, M. *Nové úpravy simulační metody Latin Hypercube Sampling a možnosti využití*. Sborník komorního semináře Problémy modelování, ISBN 80-214-2017-0, Ostrava, 16. 1. 2002.