

## I.2 Metropolis Monte-Carlo Methode

Generische Aufgabenstellung: Berechnung von Integral (oder Summe), z. B. über Teilgebiet des  $\mathbb{R}^d$

$$I = \int_V d^d r \ f(\vec{r}); \quad V \subset \mathbb{R}^d \quad \text{kompakt + endlich}$$

Deterministischer Lösungsansatz: wähle regelmäßiges Gitter, z. B. isotrop mit Schrittweite  $h$

$$\vec{r}_{\vec{n}} = h \vec{n} \quad (\vec{n} \in \mathbb{Z}^d)$$

und approximiere Integral durch diskrete Approximation (numerische Integration):

$$I \approx h^d \sum_{\substack{\vec{n} \\ \vec{r}_{\vec{n}} \in V}} f(\vec{r}_{\vec{n}})$$

**Problem:** der relative Fehler ist proportional zur Schrittweite  $h^s$  (z. B.  $h$  oder  $h^2$ ), der Aufwand  $t$  skaliert aber mit der Anzahl der Funktionsauswertungen:

$$t \approx \frac{V}{h^d} \propto h^{-d} \Rightarrow \Delta I \propto h^2 \propto t^{-2/d}$$

fällt  $\uparrow$  für große  $d$   
nur sehr langsam ab  
bei Rechenzeit  $t \rightarrow \infty$

**Beispiel 2a:** Bei Integration 2. Ordnung und  $d=200$  muss man für Halbierung des Fehlers  $2^{100} \approx 10^{30}$  mal länger rechnen.

Alternative: stochastischer Ansatz

(i) Monte Carlo (MC) ohne Gewichtung (simple Monte Carlo)

$$\Delta I \approx I_N = V \frac{1}{N} \sum_{i=1}^N f(\vec{r}_i)$$

mit zufällig, gleichverteilt gezogenen Koordinaten (d.h. konstante Wahrscheinlichkeitsdichte für mögliche Koordinaten):

$$p(\vec{r}_i) = \frac{1}{V} \quad (\text{mittels Standard-Zufallszahlengenerator})$$

Falls die Varianz  $\sigma_f^2$  von  $f$  auf  $V$  existiert (und endlich ist), konvergiert die MC-Auswertung für lange Zeiten (aut dem Gesetz der großen Zahl):

$$\Delta I_N = \frac{\sigma_f}{\sqrt{N}} \xrightarrow{N \rightarrow \infty} 0$$

$$\Delta I_N \propto t^{-\frac{1}{2}}$$

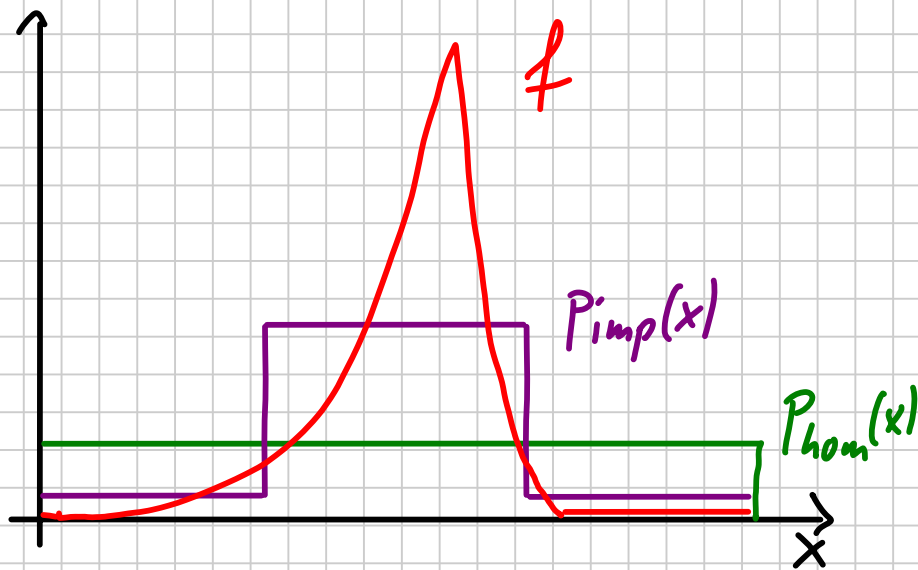
Schon bei  $d \geq 5$  konvergiert die MC-Lösung also schneller als ein Integrationsverfahren mit Fehler  $\propto h^2$ .

**Beispiel 2b:** Für Beispiel 2a ( $d=200$ ) lässt sich der Fehler (wie in allen Anwendungen) durch 4 Mal höhere Rechenzeit halbieren (statt Faktor  $10^{30}$ ).

**Problem:** die Varianz  $\sigma_f^2$  ist häufig groß, insbesondere in Anwendungen der Statistischen Physik (Schwankung des Boltzmann-Faktors  $e^{-\beta E_i}$  um viele Größenordnungen).

## (ii) Monte Carlo mit Gewichtung („importance sampling“)

Idee: konzentriere Stichproben  $\vec{r}_i$  auf Bereiche, die besonders stark zum Integral beitragen.



Man wählt also eine Verteilungsfunktion  $p(x)$ , die  $|f(x)|$  approximiert.

Formal: zerlege Integrand  $f(x)$  in Wahrscheinlichkeitsanteil  $p(x)$  und Observablenanteil  $o(x)$ :

$$I = \int d^d r f(\vec{r}) = \int d^d r p(\vec{r}) o(\vec{r})$$

$$\approx V \frac{1}{N} \sum_{i=1}^N o(\vec{r}_i); \quad \vec{r}_i \text{ zufällig mit Wahrscheinlichkeit } p(\vec{r}_i)$$

beachte: Normierung von  $p(\vec{r})$  muss bekannt sein!

$$\Delta I_N = \frac{\sigma_o}{\sqrt{N}}; \quad \sigma_o \ll \sigma_f \quad \text{für geeignete Wahl von } o(\vec{r})$$

**Problem:** Wie realisiert man inhomogene  $N^d$ -Verteilungen?

- Variablen-Transformation
- Box-Müller für Gauß-Verteilung
- Von-Neumann „rejection method“ – ineffizient
- ??? (keine allg. Methode verfügbar!)

### (iii) Markov-Ketten - Monte Carlo, Metropolis - Algorithmus

Idee: erzeuge Kette  $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n, \dots$  von Zuständen, wobei die Wahrscheinlichkeitsverteilung  $p(\vec{r}_n)$  asymptotisch gegen eine gewünschte Verteilung  $p(\vec{r})$  konvergiert.

**Definition:** Ein Stochastischer Prozess ist eine Familie von Zufallsvariablen  $(X_t)_{t \in I}$ , wobei die Indexmenge geordnet ist:  $I = \{t_0 < t_1 < \dots\}$ . Für eine abzählbare Menge  $I$  (typisch  $I = \mathbb{N}$ ) heisst der Prozess **zeitdiskret**, sonst **zeitstetig** (z. B. für  $I = \mathbb{R}_0^+$ ).

**Bemerkung:** Die Wertemengen der Zufallsvariablen können in beiden Fällen diskret oder stetig sein.

**Wichtig:** Zur vollständigen Charakterisierung eines Stochastischen Prozesses müssen neben den Wahrscheinlichkeitsverteilungen für festes  $t$  (1-Punkt-Verteilungen)

$$p_n(x_{t_n}) \equiv p(x_{t_n}, t_n)$$

i.a. auch alle  $n$ -Punkt-Verteilungen spezifiziert werden, z. B.  $p(x_{t_n}, t_n; x_{t_{n-1}}, t_{n-1}; \dots; x_{t_1}, t_1)$

**Beispiel 3a:** Wiederholter Münzwurf; bei „Zahl“ steigt der Kontostand des Spielers um 1, bei „Wappen“ sinkt er um 1.

$$p_0(x) = \delta_{x,0} \quad ; \quad p_1(x) = \frac{1}{2}(\delta_{x,-1} + \delta_{x,1}); \quad (\text{hier } t_n = n)$$

$$p_2(x) = \frac{1}{2}(\delta_{x,-2} + 2\delta_{x,0} + \delta_{x,2}) \equiv p(x_2, 2) \dots$$

**aber:**  $p(x_n, n | x_{n-1}, n-1) = \frac{1}{2}(\delta_{x_n, x_{n-1}-1} + \delta_{x_n, x_{n-1}+1})$   
↑ bedingte W'

(Allgemein gilt:  $p(x_n, n; x_m, m) = p(x_n, n | x_m, m) p(x_m, m)$ )

**Beispiel 3b:** ein modifiziertes Spiel, in dem nach jedem Wurf das Vorzeichen des Kontostandes invertiert wird, hat zwar die gleiche 1-Punkt-Verteilung wie **Bsp. 3a**, aber andere  $n$ -Punkt-Verteilungen für  $n \geq 2$ .

**Definition:** Einen stochastischen Prozess, in dem die bedingte  $W$ -Verteilung nur vom letzten bekannten Zustand abhängt, nennt man **Markov-Prozess**:

$$p(x_n, t_n | x_{i_1}, t_{i_1}, \dots, x_{i_k}, t_{i_k}) = p(x_n, t_n | x_{i_s}, t_{i_s}) \text{ für } t_{i_s} = \max_{1 \leq j \leq k} \{t_{i_j}\}$$

**Beispiel:** Zufallsbewegungen (Random Walks) wie **Bsp. 3a** sind Markov-Prozesse; auch **3b** ist Markov-Prozess.

Ein Markov-Prozess hat also ein minimales „Gedächtnis“; Prozesse ohne Gedächtnis nennt man unkorreliert.

Wichtiger Spezialfall: Markov-Prozesse mit stationären (zeitlich translationsinvarianten) Übergangswahrscheinlichkeiten

$$p(x_n, t_n | x_k, t_k) = p_t(x_n | x_k); \quad t = t_n - t_k$$