On the sparse grid combination method for SU(2) lattice problems

Michael Griebel

University of Bonn and Fraunhofer SCAI

Joint work with Heinz-Jürgen Flad, U. Bonn

1. The lattice model SU(2)
2. The sparse grid combination method
3. Numerical experiments

The lattice model SU(2)

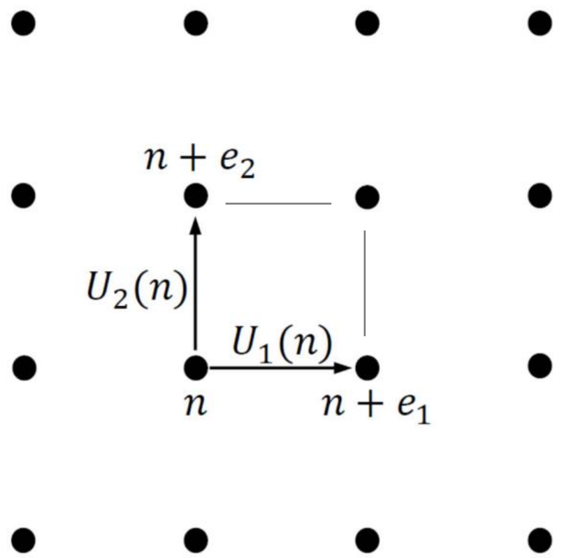
- Non-abelian SU(2) Yang-Mills theory in d dimensions
  - $d-1$ spatial and one imaginary time dimension
  - Wick-rotated into Euclidean space with Euclidean metric

- Uniform lattice of size N, i.e
$$\Lambda_{d,N} := \{n = (n_0, \ldots, n_{d-1}) \in \mathbb{N}_0^d : n_\mu = 0, \ldots, N-1\}$$

- On each lattice site $n \in \Lambda_{d,N}$, place d link variables
$$U_\mu(n) \in SU(2), \mu = 0, \ldots, d-1$$
  - Matrices $U_\mu(n) \in \mathbb{C}^{2\,x\,2}$ with $U_\mu(n) U_\nu^\dagger(n) = 1$ and $\det U_\mu(n) = 1$
  - Connect site n to sites $n + e_\mu$ in forward directions $\mu = 0, \ldots, d-1$, with e_μ the d −dimensional unit vector in coordinate direction μ.

The lattice model SU(2)

- Example



- Set of all such matrices

$$\mathcal{U}_{d,N} = \{U_\mu(n) \in SU(2), n \in \Lambda_{d,N}, \mu = 0, \dots, d-1\}$$

- Plaquette operator in site n is the closed loop

$$P_{\mu,\nu}(n) := U_\mu(n)\, U_\nu(n + e_\mu) U_\mu^\dagger(n + e_\nu) U_\nu^\dagger(n)$$

The lattice model SU(2)

- For each element in $\mathcal{U}_{d,N}$ we have Wilson's lattice action

$$S_{\Lambda_{d,N}} := -\frac{\beta}{2} \sum_{n \in \Lambda_{d,N}} \sum_{\mu<\nu} Re\ Tr\ P_{\mu,\nu}(n)$$

  with $\beta = \frac{1}{g_0^2}$ the inverse squared gauge coupling constant

- Produce a Markov chain of M samples $\mathcal{U}_{d,N,i}, i = 1, \ldots, M$ via rejection sampling by the Metropolis algorithm distributed as

$$\mathbb{P}\left(\mathcal{U}_{d,N}\right) \propto \exp(-S(\mathcal{U}_{d,N}))$$

  - Tune the rejection rate to about 50%
  - For given parameters $N, M, \beta,$ we employ the code *su2* of C. Urbach, HISKP, see

$$https://github.com/urbach/su2/tree/EducationalVersion$$

The lattice model SU(2)

- The main observable is the empirical plaquette expectation

$$\hat{P}_{d,N,M} := \langle P \rangle_M := \frac{1}{M} \sum_{i=1}^{M} P(\mathcal{U}_{d,N;i})$$

where

$$P(\mathcal{U}_{d,N;i}) = \frac{2}{d(d-1)N^d} \sum_{n \in \Lambda_{d,N}} \sum_{\mu < \nu} Re\ Tr\ U_{\mu,\nu;i}(n)$$

with the sampled matrices

$$U_{\mu,\nu:i}(n),\ \ n \in \Lambda_{d,N}, \mu, \nu = 0, \dots, d-1, \mu < \nu$$

that together comprise one sample $\mathcal{U}_{d,N,i}$

- The cost is $cost_{d,N,M} = O(MN^d)$ with β-dependent constant
- The error is $|\hat{P}_{d,\infty,\infty} - \hat{P}_{d,N,M}|$

Limits

- For $M \to \infty$ we have convergence $\hat{P}_{d,N,M} \longrightarrow \hat{P}_{d,N,\infty}$,
  - Is due to the law of large numbers and the Metropolis algorithm
- Infinite volume limit: $\mathrm{N} \to \infty$.
  - It exists for $\beta > 0$, the system converges to a Young-Mills theory in discrete space-time with lattice spacing $a = 1$
  - If the confinement property holds we have fast convergence in N and need practically only moderate values of N
- Continuum limit:
  - Hope: By making N larger while making the lattice spacing a smaller and smaller, a continuous theory is obtained with $a \to 0$.
  - Renormalization group transform, Callan-Symanzik equation, Low-Gell-Mann functions, improved actions

- For now: We stick to the infinite volume limit only

Sparse grid combination method

- So far: With d, N, M, β fixed, we can compute $\hat{P}_{d,N,M}$ by the code *su2* with cost of $O(MN^d)$, certain accuracy and convergence rate
  - Can we improve on the relation of error versus cost ?
- Sparse grids:
  - For function approximation, quadrature, PDE solution, uncertainty quantification, machine learning, we can apply the sparse grid idea and have substantial cost complexity gains
  - This is due to higher regularity of the underlying functions which possess bounded mixed derivatives.
  - Can we find a related property for lattice problems, and can we exploit it to gain faster algorithms ?
- We try to first find out for the SU(2)

- Note: Multilevel MC and multi-fidelity UQ are just special cases of a sparse grid method, sparse grids are more general

Sparse grid combination method

- Consider dyadic levels $l_1, l_2 \geq 1$, i.e. $N = 2^{l_1}, M = 2^{l_2}$ and, with

$$\tilde{P}_{l_1,l_2} := \hat{P}_{d,2^{l_1},2^{l_2}}$$

the associated table of results of the code *su2*

$$\{\tilde{P}_{l_1,l_2}\}_{l_1,l_2} \quad l_1, l_2 \in \mathbb{N}^2$$

which is trivially extended to zero levels by $\tilde{P}_{0,l_2} = \tilde{P}_{l_1,0} = \tilde{P}_{0,0} = 0$

- Define the table $\{\Delta_{l_1,l_2}\}_{l_1,l_2}$ of the hierarchical surplus/benefit

$$\Delta_{l_1,l_2} := \tilde{P}_{l_1,l_2} - \tilde{P}_{l_1-1,l_2} - \tilde{P}_{l_1,l_2-1} + \tilde{P}_{l_1-1,l_2-1}$$

- Telescopic sum identities

$$\tilde{P}_{L_1,L_2} = \sum_{l_1=1}^{L_1} \sum_{l_2}^{L_2} \Delta_{l_1,l_2} \quad \text{and} \quad \tilde{P}_{\infty,\infty} = \sum_{l_1=1}^{\infty} \sum_{l_2}^{\infty} \Delta_{l_1,l_2}$$

Sparse grid combination method

- Take the cost $cost_{l_1,l_2}$ of each Δ_{l_1,l_2} into account, which is up to a small constant that of the cost of $\tilde{P}_{l_1,l_2}$
- Define the table $\{bcr_{l_1,l_2}\}_{l_1,l_2}$ with the benefit/cost ratios

$$bcr_{l_1,l_2} := \frac{|\Delta_{l_1,l_2}|}{cost_{l_1,l_2}}$$

- The optimal index set Γ_K can be determined by a simple knapsack problem: Sort the benefit/cost ratios and take the first K indices with the largest bcr_{l_1,l_2} into account
- For rising K it involves a truncation of the *bcr* table along its level set lines
- Leads to the associated general sparse grid approximation

$$\tilde{P}_{L_1,L_2} = \sum_{l_1,l_2 \in \Gamma_K} \Delta_{l_1,l_2}$$

with minimal overall cost $\sum_{l_1,l_2 \in \Gamma_K} cost_{l_1,l_2}$ and minimal error

Sparse grid combination method

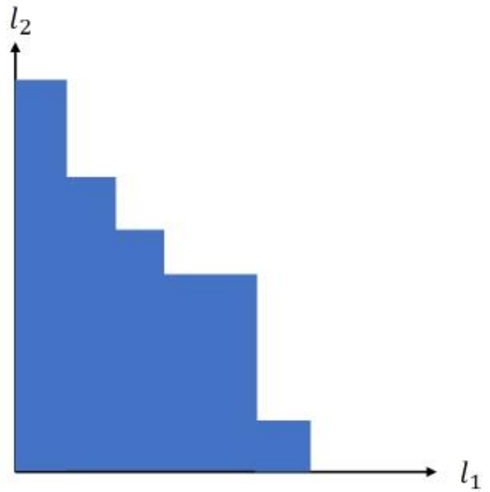
- The telescopic sum for Γ_K can be partially reversed. This leads to the general sparse grid combination method

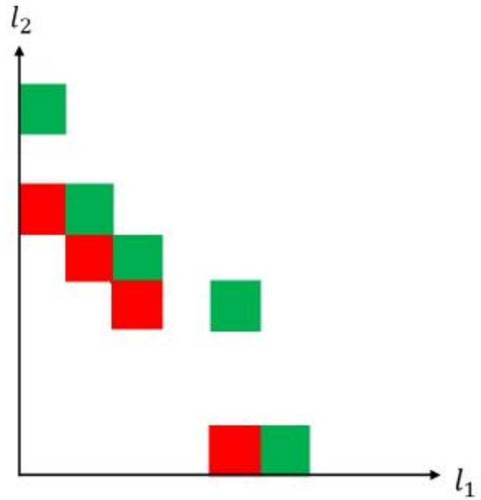
$$\tilde{P}_{\Gamma_K,} = \sum_{l_1,l_2\in \Gamma_K} \Delta_{l_1,l_2} = \sum_{l_1,l_2\in \Gamma_K} c_{l_1,l_2\,l_1,l_2} \tilde{P}_{l_1,l_2}$$

with the combination coefficients

$$c_{l_1,l_2} \coloneqq \sum_{z_1,z_2=0,0}^{l_1,l_2} (-1)^{z_1+z_2} \chi_{\Gamma_K}((l_1,l_2)+(z_1,z_2))$$

and the characteristic function $\chi_{\Gamma_K}(l_1,l_2) \coloneqq \begin{cases} 1 & \text{if } (l_1,l_2) \epsilon\, \Gamma_K \\ 0 & \text{else} \end{cases}$





Sparse grid combination method

- Involves now only certain $\tilde{P}_{l_1,l_2}$, i.e. calls of the code *su2*, with different parameters and the linear combination of its results

- Can be tried analogously with other theories and codes

- Can also be tried for code involving renormalization, Callan-Symanzik corrections, etc. provided that there is code and that the limit exists at all ?

- Can be seen as a two-variate extrapolation method between the lattice size/spacing and the number of MC samples

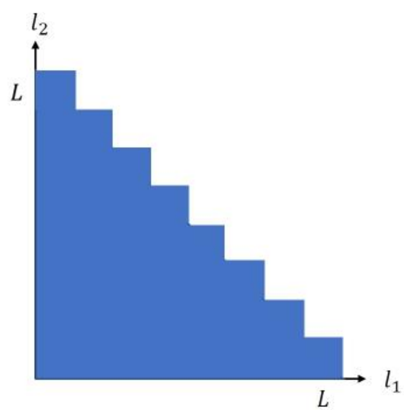
Sparse grid combination method

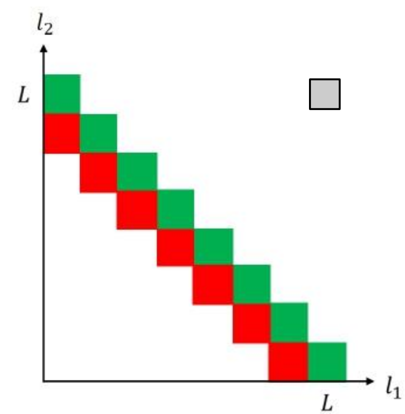
- Most simple example: Isotropic sparse grid, $d = 2$
- Index set: $\Gamma_{isp,L} = \{(l_1, l_2) \epsilon \mathbb{N}^2, l_1 + l_2 \leq L + 1 \}$
- Isotropic sparse grid approximation

$$\tilde{P}_{isp,L} = \sum_{l_1,l_2 \in \Gamma_{sparse,L}} \Delta_{l_1,l_2}$$

- Isotropic combination technique

$$\tilde{P}_{isp,L} = \sum_{l_1+l_2=L+1} \tilde{P}_{l_1,l_2} - \sum_{l_1+l_2=L+1} \tilde{P}_{l_1,l_2}$$





- Cost complexity gain in contrast to full grid ☐:

$O(L2^{L})$ instead of $O(2^{2L})$

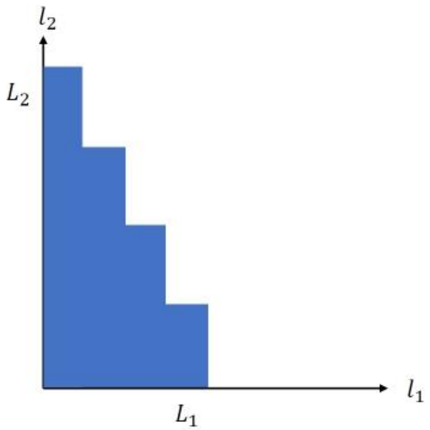
Sparse grid combination method

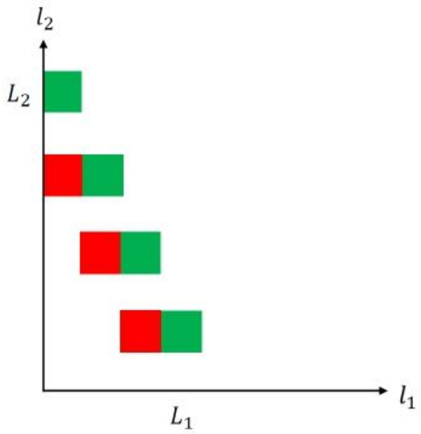
- Isotropic situation rarely encountered: Slow rate of MC in l_1-direction, much faster rate in l_2-direction, product decay unclear
- Anisotropic index set associated to L_1, L_2

$$\Gamma_{asp,L_1,L_2} = \{(l_1,l_2)\epsilon\mathbb{N}^2, (L_2-1)l_1+(L_1-1)l_2 \leq L_1L_2-1\}$$

- Anisotropic sparse grid approximation $\tilde{P}_{asp,L_1,L_2} = \sum_{l_1,l_2\in\Gamma_{asp,L_1,L_2}} \Delta_{l_1,l_2}$
- Anisotropic combination technique

$$\tilde{P}_{asp,L_1,L_2} = \sum_{(L_2-1)l_1+(L_1-1)l_2= L_1L_2} \tilde{P}_{l_1,l_2} - \sum_{(L_2-1)l_1+(L_1-1)l_2== L_1L_2-1} \tilde{P}_{l_1,l_2}$$





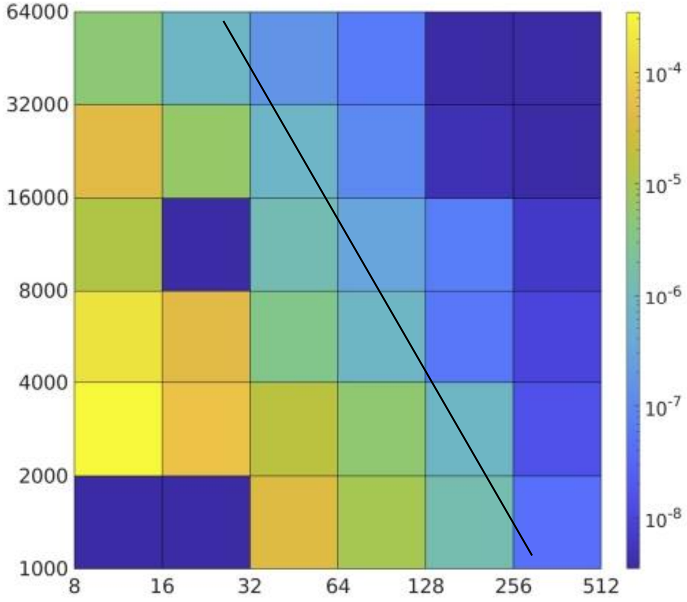
- Cost complexity gain in contrast to full grid:

$$O(2^{\max(L_1,L_2)}) \text{ instead of } O(2^{L_1+L_2})$$

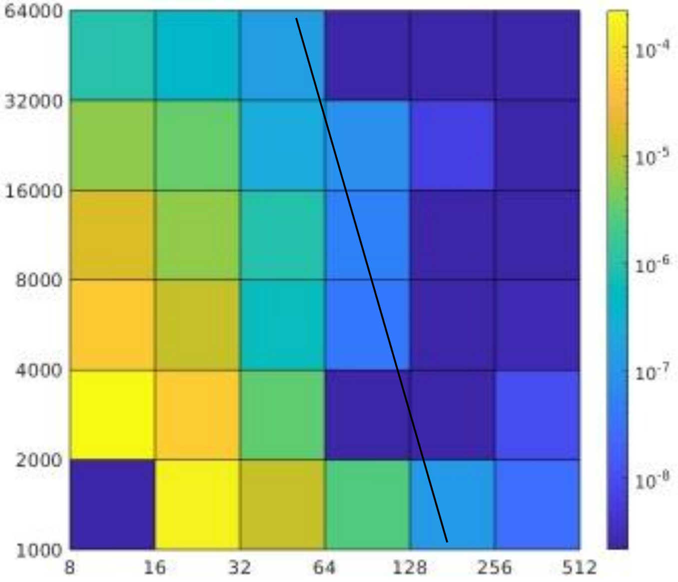
Numerical experiments

- For simplicity, set $d = 2$
- We consider $\beta = 0{,}5$ and $\beta = 10$

- Benefit cost ratios for N, M of different level indices



$\beta = 0{,}5$



$\beta = 10.0$

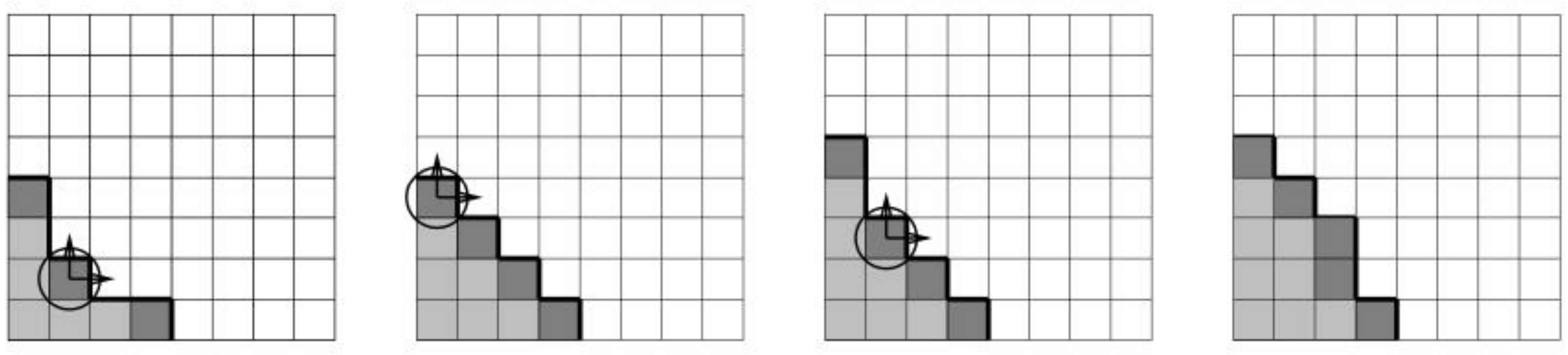
- We see an anisotropic sparse grid structure in both cases and NOT a full grid structure

Numerical experiments

- This is good news:
  - We have a kind of product type behavior of the convergence and a sparse grid effect for the SU(2).
  - Allows substantially faster algorithms by means of the sparse grid combination method
- The diagonal cut off isoline of the isotropic case for optimal complexity is rotated
  - Anisotropy reflects the slower convergence rate in sampling direction versus the faster rate in lattice size direction
  - Gets more profound for the larger value of β
- How can we detect this algorithmically and how can we practically construct an optimal index set ?

Algorithmic remarks

- Dimension-adaptive method builds the index set successively adapted to a specific problem under consideration
- Example:



  - Adaptive method: Error indicator $bcr_{l_1,l_2} > \varepsilon$ involving the Δ_{l_1,l_2}, largest value marks the index for refinement, refinement in two directions, and repeat
  - Associated combination method involves only the $\tilde{P}_{l_1,l_2}$
- Note again: Just code to be called for different lattice resolutions and different samplings/chain lengths

Concluding remarks

- We studied SU(2) lattice problems for the most simple case d=2 and the infinite volume limit
  - Wanted to find out if there is a kind of product decay/sparse grid effect or not
  - Yes ! Allows to substantially speed up calculations

- Next:
  - Consider the cases $d = 3, 4$
  - Try to consider the continuum limit case, code ?
  - Consider other problems than just SU(2), code ?
  - Note: Other codes can be simply plugged into our method

- Instead with MCMC, our approach could work for quasi MC-type techniques of higher order as well ?