A Preconditioner for Improved Fermion Actions

N. Eicker\textsuperscript{a}, W. Bietenholz\textsuperscript{a}, A. Frommer\textsuperscript{b}, Th. Lippert\textsuperscript{c}, B. Medeke\textsuperscript{b}, and K. Schilling\textsuperscript{a,c}

\textsuperscript{a}HLRZ, c/o Research Center Jülich, D-52425 Jülich, Germany, and DESY/Hamburg, Germany
\textsuperscript{b}Department of Mathematics, University of Wuppertal, D-42097 Wuppertal, Germany
\textsuperscript{c}Department of Physics, University of Wuppertal, D-42097 Wuppertal, Germany

SSOR preconditioning of fermion matrix inversions which is parallelized using a \textit{locally-lexicographic} lattice sub-division has been shown to be very efficient for standard Wilson fermions. We demonstrate here the power of this method for the Sheikholeslami-Wohlert improved fermion action and for a renormalization group improved action incorporating couplings of the lattice fermion fields up to the diagonal in the unit hypercube.

1. Introduction

Recently, the symmetric successive over-relaxed preconditioner (SSOR) turned out to be parallelizable by means of the \textit{locally-lexicographic} ordering technique \cite{1}. In this way, SSOR preconditioning has been made applicable to the acceleration of standard Wilson fermion inversions on high performance massively parallel systems and it outperforms o/e preconditioning.

It appears intriguing to extend the range of \textit{ll}-SSOR-preconditioners such as to accelerate the inversion of improved fermionic actions, which became very popular in the recent years.

In Symanzik’s \textit{on-shell improvement} program \cite{2}, counter terms are added to both, lattice action and composite operators in order to reduce $O(a)$ artifacts which spoil results in the instance of the Wilson fermion formulation. In the approach of Sheikholeslami and Wohlert (SWA) \cite{3}, the Wilson action is modified by adding a diagonal term, the so-called clover term with a new free parameter $c_{SW}$.

\textit{Perfect lattice actions} are situated on renormalized trajectories in parameter space that intersect the critical surface (at infinite correlation length) in a fixed point of a renormalization group transformation. Perfect actions are free of any cutoff effects, but in practice they can only be constructed approximatively. A promising approach for asymptotically free theories is the use of classically perfect actions \cite{4} to serve as an approximation to perfect ones. Moreover, practical applications require a truncation of the couplings to short distances (truncated perfect actions, TPA).

In the present investigation, we consider a variant of the hypercube fermion (HF) approximation \cite{5}. The generic form of both SWA and TPA is given by

\begin{equation}
M = D + A + B + C + E \ldots
\end{equation}

$D$ stands for 12 $\times$ 12 diagonal sub-blocks, $A, B, \ldots$ are nearest-neighbor, next-to-nearest-neighbor, $\ldots$ hopping terms. In the following, we will show that the \textit{ll}-SSOR scheme applies not only to the couplings in $A$ but also to the internal spin and colour d.o.f. of $D$ (SWA) as well as all the couplings of $B, C,$ and $E, \ldots$ of TPA.

2. SWA and HF Actions

SWA is composed of $A$ (Wilson hopping term) and $D$ (SW diagonal):

\begin{equation}
D_{SW}(x,y) = \left[1 + \frac{c_{SW}}{2} \kappa \sum_{\mu,\nu} \sigma_{\mu\nu} F_{\mu\nu}(x)\right] \delta_{x,y};
\end{equation}

\begin{equation}
A_{SW}(x,y) = -\kappa \sum_{\mu} \left[ (1 - \gamma_{\mu}) U_{\mu}(x) \delta_{x,y-\mu} + (1 + \gamma_{\mu}) U_{-\mu}(x) \delta_{x,y+\mu}\right],
\end{equation}

$\kappa$ is the Wilson hopping parameter, $c_{SW}$ couples the SW clover operator. This parameter is tuned

\textsuperscript{*}Talk presented by N. Eicker
to optimize $O(a)$ cancellations. The clover term $F_{\mu}(x)$ consists of $12 \times 12$ diagonal blocks. Its explicit structure in Dirac space is given in Ref. [1].

As a prototype TPA we have investigated the perfect free action constructed in Ref. [7] by means of block variable renormalization group transformations for free fermions. The exponential decay of their couplings is fast, and therefore they can be truncated to short range [5]. We limit ourselves to couplings up to 4-space diagonals in the unit hypercube (hypercube fermion, HF). The gauge links are introduced in an obvious way: we connect all the coupled sites by all possible $d!$ shortest lattice paths in a $d$ diagonal, by multiplying the compact gauge fields on the path links. For a given link, we average over all $d!$ paths from hyper-links $U^{(d)}_{\mu_1+\mu_2+\ldots+\mu_d}(x)$ built up recursively:

$$U^{(d)}_{\mu_1+\mu_2+\ldots+\mu_d}(x) = \frac{1}{d!} \left[ U^{(1)}_{\mu_1}(x) U^{(d-1)}_{\mu_2+\mu_3+\ldots+\mu_d}(x + \hat{\mu}_1) + U^{(1)}_{\mu_2}(x) U^{(d-1)}_{\mu_1+\mu_3+\ldots+\mu_d}(x + \hat{\mu}_2) + \ldots + U^{(1)}_{\mu_d}(x) U^{(d-1)}_{\mu_1+\mu_2+\ldots+\mu_{d-1}}(x + \hat{\mu}_d) \right].$$

(3)

Defining effective $\Gamma$'s by

$$\Gamma_{\pm \mu_1 \pm \mu_2 \pm \ldots \pm \mu_d} = \lambda_d + \kappa_d (\pm \gamma_{\mu_1} \pm \gamma_{\mu_2} \pm \ldots \gamma_{\mu_d}),$$

(4)

with the HF hopping parameters $\kappa_i$ and $\lambda_i$, we arrive at:

$$D_{HF}(x, y) = \lambda_0 \delta_{x,y},$$

(5)

$$A_{HF}(x, y) = \sum_{\mu} \left[ \Gamma_{+\mu} U^{(1)}_{\mu}(x) \delta_{x,y-\hat{\mu}} + \Gamma_{-\mu} U^{(1)}_{-\mu}(x) \delta_{x,y+\hat{\mu}} \right],$$

$$B_{HF}(x, y) = \sum_{\mu} \left[ \Gamma_{\mu+\nu} U^{(2)}_{\mu+\nu}(x) \delta_{x,y-\hat{\nu}} + \Gamma_{\mu-\nu} U^{(2)}_{\mu-\nu}(x) \delta_{x,y+\hat{\nu}} + \Gamma_{\mu+\nu} U^{(2)}_{\mu+\nu}(x) \delta_{x,y+\hat{\nu}} + \Gamma_{-\mu-\nu} U^{(2)}_{-\mu-\nu}(x) \delta_{x,y+\hat{\nu}} \right].$$

It is straightforward to write down the expressions for $C_{HF}$ and $E_{HF}$. Altogether 80 hyper-links contribute while 40 have to be stored.

### 3. Block SSOR Preconditioning

The preconditioned system is modified by two matrices $V_1$ and $V_2$:

$$V_1^{-1}MV_2^{-1}\vec{\psi} = \vec{\phi}, \quad \vec{\phi} := V_1^{-1}\vec{\phi}, \quad \vec{\psi} := V_2\vec{\psi}. \quad (6)$$

Let $M = D - L - U$ be the decomposition of $M$ into its block diagonal part $D$, its (block) lower triangular part $-L$ and its (block) upper triangular part $-U$. Block SSOR preconditioning is defined through the choice

$$V_1 = \left( \frac{1}{\omega} D - L \right) \left( \frac{1}{\omega} D \right)^{-1}, \quad V_2 = \frac{1}{\omega} D - U. \quad (7)$$

The Eisenstat trick [6] reduces the costs by a factor 2. It is based on the identity:

$$V_1^{-1}(D - L - U) V_2^{-1} = (1 - \omega LD^{-1})^{-1} \times \left[ 1 + (\omega - 2) \left( 1 - \omega UD^{-1} \right)^{-1} \right] + (1 - \omega UD^{-1})^{-1}. \quad (8)$$

The preconditioned matrix-vector product, $z = V_1^{-1}MV_2^{-1}x$, is given by:

- **solve** $(1 - \omega UD^{-1})y = x$
- **compute** $w = x + (\omega - 2)y$
- **solve** $(1 - \omega LD^{-1})v = w$
- **compute** $z = v + y$

The "solve" is just a simple forward (backward) substitution process due to the triangular structure:

for $i = 1$ to $N$

$$v_i = v_i + \sum_{j=1}^{i-1} L_{ij}s_j$$

$$s_i = \omega D_{ii}^{-1}v_i$$

Options for $D$ of SWA take each block $D_{ii}$ to be of dimension 12 ($D^{(12)}$), 6 ($D^{(6)}$), 3 ($D^{(3)}$) or 1 ($D^{(1)}$), as suggested by the structure of $D$. The blocks have to be pre-inverted the costs depending on the block size [3].

Parallelism can be achieved by **locally lexicographic ordering** [3]. “Coloring” is the decomposition of all lattice points into mutually disjoint sets $C_1, \ldots, C_k$ (with respect to the matrix $M$), if for any $l \in \{1, \ldots, k\}$ the property $x \in C_l \Rightarrow y \not\in C_l$ for all $y \in n(x)$ holds. $n(x)$ denotes the set of sites $\neq x$ coupled to $x$. A suitable ordering first numbers all $x$ with color $C_1$, then all with $C_2$ etc. Thus, each lattice
point couples with lattice points of different colors only. The computation of $v_x$ for all $x$ of a given color $C_1$ can be done in parallel, since terms like $\sum_{y \in n(x), y \leq x}^{}$ involve only lattice points of the preceding colors $C_1, \ldots, C_{l-1}$, with $x \leq y$ meaning that $x$ has been numbered before $y$ with respect to the ordering $o$.

Let the lattice blocks be of size $n^{loc} = n^{loc}_1 \times n^{loc}_2 \times n^{loc}_3 \times n^{loc}_4$. A different color is associated with each of the sites of the $n^{loc}$ groups. A locally lexicographic (ll) ordering is defined to be the color ordering, where all points of a given color are ordered after all points with colors, which correspond to lattice positions on the local grid that are lexicographically preceding the given color.

The parallel forward substitution reads:

\[
\text{for } C_i, i = 1, \ldots, n^{loc}, \text{ fracnp } \in \mathbb{N} \\
\quad \text{for all processors } j = 1, \ldots, p \\
\quad \quad x := \text{ with } C_i \text{ on } j \\
\quad \quad v_x = w_x + \sum_{y \in n(x), y \leq x}^{} L_{xy} s_y \\
\quad \quad s_x = \omega D^{-1} v_x
\]

For SWA, up to 8 and for HF all 80 neighbors may be involved on the 4-d grid.

4. Improvement

The SWA has been implemented on an APE100. HF is benchmarked on a DEC alpha workstation. For SWA, we use a de-correlated set of 10 quenched gauge configurations generated on a $16^4$ lattice at $\beta = 6.0$ at 3 values of $c_{SW}$, 0, 1.0 and 1.769. We have applied BiCGStab as iterative solver. The stopping criterion has been chosen as $\frac{||X - \phi||}{||X||} \leq 10^{-6}$, with $X$ being the solution. We used a local source $\phi$ and determined the optimal OR parameter to be about $\omega = 1.4$ for all block sizes and $c_{SW}$.

We plot the ratio of iteration numbers of the odd-even procedure vs. ll-SSOR as function of $\kappa$ in Fig. 1. A gain factor up to 2.5 in iteration numbers can be found. There is no dependence on $c_{SW}$ or on the block size of $D$ and only 10% on the local lattice size. As to real CPU costs on APE100, the optimal block size of $D$ is a $3 \times 3$ block whereas on a scalar system, the optimum is found for a $1 \times 1$ diagonal.

Limited by the number of hyper-links to store on the DEC system, we decided to investigate HF on a lattice of size $8^4$. We measured at $\beta = 6.0$ in quenched QCD. We have assessed the critical mass parameter to determine the critical region of HF. For HF $\omega \simeq 1$ is optimal. We find that SSOR preconditioning of HF leads to gain factors $\simeq 3$ close to the critical bare mass $m_c = -0.92$.

REFERENCES

1. S. Fischer et al.: Comp. Phys. Comm. 98 (1996) 20.
2. K. Symanzik: Nucl. Phys. B212 (1983) 1.
3. B. Sheikholeslami and R. Wohlert: Nucl. Phys. B259 (1985) 572.
4. P. Hasenfratz and F. Niedermayer: Nucl. Phys. B414 (1994) 785.
5. W. Bietenholz et al.: Nucl. Phys. B (Proc. Suppl.) 53 (1997) 921.
6. N. Eicker et al.: hep-lat/9807013
7. W. Bietenholz and U.-J. Wiese: Nucl. Phys. B464 (1996) 319.