CUDA Implementation of CG Inverters for the Faddeev-Popov Matrix

Attilio Cucchieri[1], *Tereza Mendes*[1]

[1]Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos, SP, Brazil

DOI: http://dx.doi.org/10.3204/DESY-PROC-2014-05/38

The **strong force** is one of the four fundamental interactions of nature (along with gravity, electromagnetism and the weak nuclear force). It is the force that holds together protons and neutrons in the atomic nucleus. The strong interaction is described by Quantum Chromodynamics (QCD) [1], a quantum field theory with local gauge invariance, given by the $SU(3)$ group symmetry. A unique feature of the strong force is that the particles that feel it directly (quarks and gluons) are completely hidden from us, i.e. they are never observed as free particles. This property is known as **color confinement** and makes QCD much harder to handle than the theories describing the weak and electromagnetic forces. Indeed, it is not possible to study QCD analytically in the limit of small energies, or large spatial separations, which corresponds to several processes of interest, including the mechanism of color confinement. To gain insight into these issues, physicists must rely on numerical simulations performed on supercomputers. These studies are carried out using the **lattice** formulation of QCD [2, 3, 4, 5, 6]. Similar studies are done also for other non-Abelian $SU(N)$ gauge theories.

A **propagator** of a field is a two-point function, i.e. a correlation function between values of the field at two different points in space-time [7]. In quantum mechanics, the propagator determines the evolution of the wave function of a system and, for a particle, it gives the probability amplitude of propagating from a point in space-time to another [8]. More generally, **Green's functions** (i.e. n-point functions) carry all the information about the physical and mathematical structure of a quantum field theory. Thus, the study of the long-range —or infrared (IR)— behavior of propagators and vertices is an important step in our understanding of QCD. In particular, the confinement mechanism for color charges could manifest itself in the IR behavior of (some of) these Green's functions.

For gauge theories, such as QCD, the local gauge invariance implies that Green's functions are usually gauge-dependent quantities and can be evaluated only after a specific gauge condition is imposed. Among the possible choices, the so-called Landau (or Lorenz) gauge condition is particularly interesting, since it preserves the relativistic covariance of the theory. The status of lattice studies of IR propagators in Landau gauge has been reviewed in [9].

On the lattice, the **minimal Landau gauge** condition is usually implemented by (numerically) finding local minima of a functional [10]. As a consequence, in this gauge, the path integral over gauge-field configurations is restricted to the set of transverse configurations for which the so-called Landau-gauge **Faddeev-Popov matrix** (FP) is semi-positive-definite [11].

Thus, this matrix should encode all the relevant (non-perturbative) aspects of the theory, related to the color-confinement mechanism.

For a given (thermalized and gauge-fixed) lattice gauge-field configuration $U_\mu(x) \in SU(N)$, with $\mu = 0, 1, 2$ and 3, the FP matrix $\mathcal{M}_U$ in minimal Landau gauge is defined by its action on a function $\gamma^c(x)$ as (see, for example, [12] and [13])

$$\begin{aligned}\left(\mathcal{M}_U\gamma\right)^b(x) \;=\; \sum_\mu \Gamma^{bc}_\mu(x)\,[\,\gamma^c(x) \,-\, \gamma^c(x+e_\mu)\,] \,+\, \Gamma^{bc}_\mu(x-e_\mu)\,[\,\gamma^c(x) \,-\, \gamma^c(x-e_\mu)\,] \\ +f^{bdc}\,\left[\,A^d_\mu(x)\,\gamma^c(x+e_\mu) \,-\, A^d_\mu(x-e_\mu)\,\gamma^c(x-e_\mu)\,\right]\,.\end{aligned}$$

Here, $b, c, d = 1, \ldots, N^2-1$ are color indices, f^{bcd} are the (anti-symmetric) structure constants of the $SU(N)$ gauge group, $A^d_\mu(x)$ are the gauge fields defined by the relation

$$A^d_\mu(x)\,\lambda^d \;=\; \left.\frac{U_\mu(x) - U^\dagger_\mu(x)}{2iag_0}\right|_{\text{traceless}}\,,$$

where λ^d are the generators of the $SU(N)$ group, g_0 is the bare coupling constant, a is the lattice spacing and

$$\Gamma^{bc}_\mu(x) \;=\; \frac{1}{8}\;\mathrm{Tr}\,\left(\,\{\,\lambda^b,\lambda^c\,\}\,\left[U_\mu(x) + U^\dagger_\mu(x)\right]\right)\,.$$

We note that, in the $SU(2)$ case, one finds $\Gamma^{bc}_\mu(x) = \delta^{bc}\,\mathrm{Tr}\,U_\mu(x)/2$ and $f^{bcd} = \epsilon^{bcd}$, where ϵ^{bcd} is the completely anti-symmetric tensor. Also, note that the FP matrix $\mathcal{M}_U$ becomes the lattice Laplacian if $\Gamma^{bc}_\mu(x) = \delta^{bc}$ and $A^d_\mu(x) = 0$.

The **inverse** of the FP matrix enters the evaluation of several fundamental Green's functions of the theory, such as the ghost propagator, the ghost-gluon vertex, the Bose-ghost propagator, etc. These functions can be computed on the lattice through **Monte Carlo simulations** [14]. However, the numerical inversion of the FP matrix is rather time consuming, since it is a very large (sparse) matrix with an extremely small (positive) eigenvalue, thus requiring the use of a **parallel preconditioned conjugate-gradient** (CG) algorithm [15]. Moreover, this inversion has to be done in double precision and, for each lattice configuration, one has to consider hundreds of different sources, corresponding to different kinematic combinations. One should also stress that, in a lattice simulation, one cannot study momenta smaller than $2\pi/L$, where L is the lattice side. Thus, numerical studies of Green's functions in the IR limit (small momenta) require very large lattice volumes and a careful extrapolation of the data to the infinite-volume limit. In fact, inversion of the FP matrix is the performance bottleneck for these numerical studies.

In this study we considered four preconditioned conjugate-gradient algorithms. In particular, for the preconditioner matrix $\mathcal{P}$ we used:

i) the diagonal elements (with respect to color and space-time indices) of the FP matrix,

ii) the diagonal elements (with respect to space-time indices only) of the FP matrix,

iii) the usual lattice Laplacian [16, 17],

iv) the FP matrix with $A^d_\mu(x) = 0$.

In the former two cases described above, the inversion of $\mathcal{P}$ can be done exactly and it does not require inter-GPU communication. On the other hand, for the latter two choices, we employed a (non-preconditioned) CG algorithm. We tested the above choices for the preconditioning step using double and single precision. For the last two cases (choices **iii** and **iv**) we also considered two different stopping criteria for the CG algorithm used to invert the preconditioner matrix $\mathcal{P}$. The code has been written using `CUDA` and `MPI` and tested on multiple GPUs (Tesla S1070 and Tesla K20) interconnected by InfiniBand.

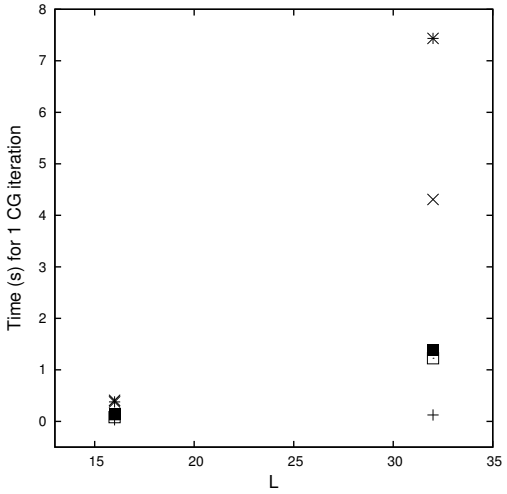

Figure 1: Weak scaling using lattice volumes $V = L^4 = 16^4$ and 32^4 with 1 and 16 Tesla S1070 GPUs, respectively. The data correspond to the CG algorithm without preconditioning (+), with preconditioning **iii** (×), **iv** (∗), **iii** in single precision (□) and **iv** in single precision (■).

Our results can be seen in the three plots in Figs. 1 and 2. In particular, we show the processing time for one CG iteration as a function of the lattice side for a fixed *lattice volume/number of GPUs* ratio (weak scaling, Fig. 1) and as a function of the number of GPUs at a fixed lattice size L (strong scaling, Fig. 2). One clearly sees that the overhead due to the inversion of the preconditioner matrix $\mathcal{P}$ is quite large for the cases **iii** and **iv** (see Fig. 1). However, this overhead can be drastically reduced by applying the preconditioner step in single precision (see again Fig. 1). Moreover, this reduction in the processing time for one CG iteration does not affect the convergence properties of the method, i.e. the number of CG iterations necessary to satisfy a given convergence criterion is essentially unchanged when moving from double to single precision. From Fig. 2 we also see that the overhead due to inter-GPU communication is not particularly critical for the algorithms considered, even when the inversion of the preconditioner matrix $\mathcal{P}$ requires inter-GPU communication and for a rather small lattice volume, at least when going from 1 to 4 GPUs.

Of the four preconditioners considered here, the last two are quite effective in reducing the number of CG iterations, typically by a factor of 3–4. We plan to study other preconditioners for the FP matrix, including the so-called even-odd preconditioning, and to fine-tune the

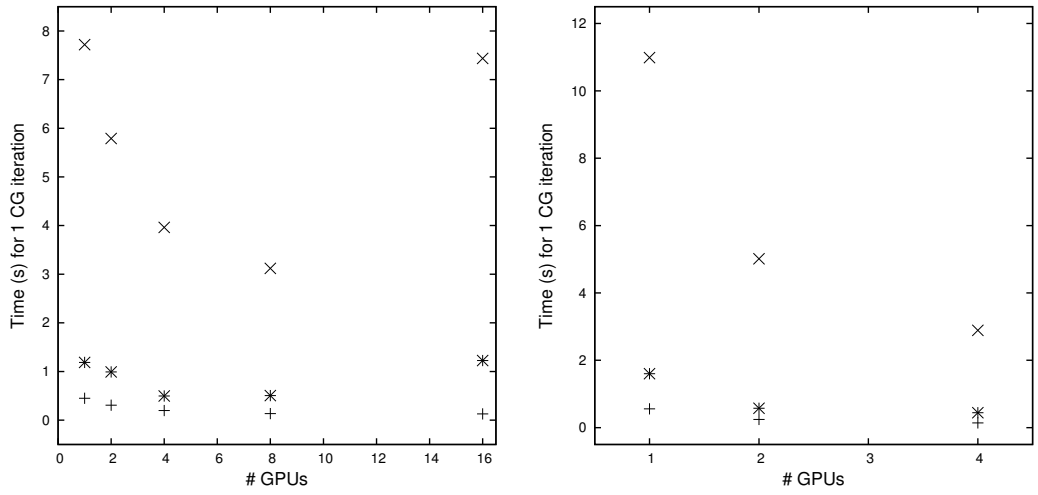

Figure 2: Strong scaling using lattice volume $V = 32^4$ with, respectively, 1,2,4,8,16 Tesla GPUs (left plot) and 1,2,4 Kepler GPUs (right plot). The data correspond to the CG algorithm with preconditioning **i** (+), with preconditioning **iv** (×) and with preconditioning **iii** in single precision (∗).

implementation of the various algorithms.

We acknowledge FAPESP, PRP–USP and CNPq for financial support.

References

[1] *An Elementary Primer for Gauge Theory*, K. Moriyasu, (World Scientific, Cingapura, 1983).

[2] *Quarks, gluons and lattices*, M. Creutz, (Cambridge University Press, Cambridge UK, 1983).

[3] *Introduction to Quantum Fields on a Lattice*, J. Smit, (Cambridge University Press, Cambridge UK, 2002).

[4] *Lattice gauge theories. An introduction*, H.J. Rothe, (World Scientific, Cingapura, 2005).

[5] *Lattice Methods for Quantum Chromodynamics*, T. DeGrand e C. DeTar, (World Scientific, Cingapura, 2006).

[6] *Quantum Chromodynamics on the Lattice*, C. Gattringer e C.B. Lang, (Springer, Berlim D, 2010).

[7] *Quantum and Statistical Field Theory*, M. Le Bellac, (Oxford University Press, Oxford UK, 1995).

[8] *Modern Quantum Mechanics*, J.J. Sakurai, edited by San Fu Tuan, (revised edition, Addison-Wesley Publishing Company, Reading MA, USA, 1994).

[9] A. Cucchieri and T. Mendes, PoS **QCD-TNT09** 026 (2009).

[10] L. Giusti, M.L. Paciello, C. Parrinello, S. Petrarca and B. Taglienti, Int. J. Mod. Phys. **A16** 3487 (2001).

[11] N. Vandersickel and D. Zwanziger, Phys. Rept. **520** 175 (2012).

[12] D. Zwanziger, Nucl. Phys. **B412** 657 (1994).

[13] A. Cucchieri, T. Mendes and A. Mihara, Phys. Rev. **D72** 094505 (2005).

[14] *A guide to Monte Carlo simulations in Statistical Physics*, D.P. Landau e K. Binder, (Cambridge University Press, Cambridge UK, 2000).

[15] *Iterative methods for sparse linear systems*, Y. Saad, (2nd edition, SIAM, Philadelphia, PA, 2003).

[16] S. Furui and H. Nakajima, Phys. Rev. **D69** 074505 (2004).

[17] A. Sternbeck, E.-M. Ilgenfritz, M. Muller-Preussker and A. Schiller, Phys. Rev. **D72** 014507 (2005).