CUDA Implementation of CG Inverters for the Faddeev-Popov Matrix

Attilio Cucchieri & Tereza Mendes

Instituto de Física de São Carlos – Universidade de São Paulo, Brazil

1. QCD and Color Confinement

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), a quantum field theory with local gauge symmetry, given by the $SU(3)$ group.

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. Similar studies are done also for other non-Abelian $SU(N)$ gauge theories.

2. Confinement in Landau gauge

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. 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. 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.

3. The Faddeev-Popov matrix

On the lattice, the (**minimal**) **Landau gauge** condition is usually implemented by (numerically) finding local minima of a functional. 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. Thus, this matrix should encode all the relevant (non-perturbative) aspects of the theory, related to the color-confinement mechanism. In particular, 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 through **Monte Carlo** simulations using the lattice formulation of gauge theories. However, the numerical inversion of the FP matrix is rather time consuming, since it is a huge (sparse) matrix with an extremely small (positive) eigenvalue, thus requiring the use of a **parallel preconditioned conjugate-gradient** (CG) algorithm. Moreover, for each lattice configuration, this inversion has to be done for hundreds of 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.

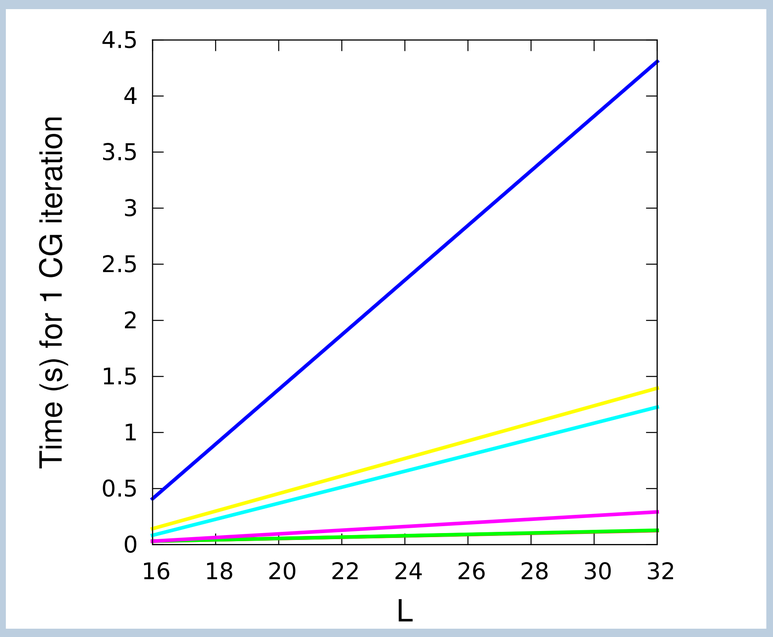
For a given (thermalized and gauge-fixed) lattice configuration $U_\mu(x) \in SU(N)$, with $\mu = 0,1,2,3$, the FP matrix $\mathcal{M}_U$ in minimal Landau gauge is defined as

$$\left(\mathcal{M}_U \gamma\right)^b(x) = \sum_\mu \Gamma^{bc}_\mu(x)\left[\gamma^c(x) - \gamma^c(x+e_\mu)\right] + \Gamma^{bc}_\mu(x-e_\mu)\left[\gamma^c(x) - \gamma^c(x-e_\mu)\right] + f^{bdc}\left[A^d_\mu(x)\gamma^c(x+e_\mu) - A^d_\mu(x-e_\mu)\gamma^c(x-e_\mu)\right]. \quad (1)$$

Here, f^{bcd} are the (anti-symmetric) structure constants of the $SU(N)$ gauge group, $A^d_\mu(x)$ are the gauge fields [related to $U_\mu(x)$] and $\Gamma^{bc}_\mu(x) = \mathrm{Tr}\left(\left\{\lambda^b,\lambda^c\right\}\left[U_\mu(x) + U^\dagger_\mu(x)\right]\right)/8$. In the $SU(2)$ case, one finds $\Gamma^{bc}_\mu(x) = \delta^{bc}\,\mathrm{Tr}\,U_\mu(x)/2$ and $f^{bcd} = \varepsilon^{bcd}$, where ε^{bcd} is the completely anti-symmetric tensor.

4. Conjugate-Gradient Inversion

In this study we considered several 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 [note that the FP matrix becomes the lattice Laplacian if $\Gamma^{bc}_\mu(x) = \delta^{bc}$ and $A^d_\mu(x) = 0$], **iv)** the FP matrix with $A^d_\mu(x) = 0$. While in the former two cases the inversion of $\mathcal{P}$ can be done exactly and it does not require inter-GPU communication, for the latter two choices we employed a (non-preconditioned) CG algorithm.



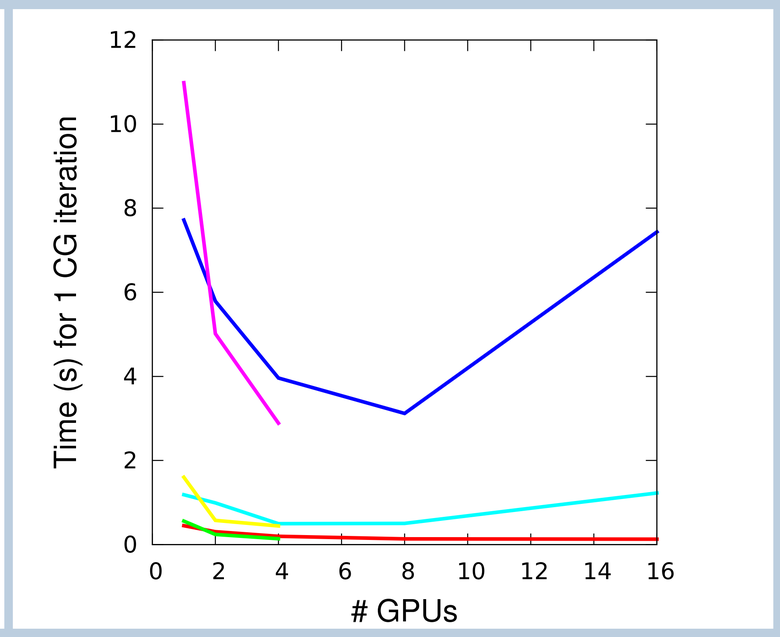


Figure 1 (left): weak scaling using lattice volumes 16^4 and 32^4 with 1 and 16 Tesla GPUs, respectively. The six lines correspond to the CG algorithm: without preconditioning (red line), with preconditioning **i** (green line), **iii** (blue line), **i** in single precision (magenta line), **iii** in single precision (cyan line) and **iv** in single precision (yellow line). **Figure 2 (right):** strong scaling using lattice volume 32^4 with, respectively, 1,2,4,8,16 Tesla GPUs and 1,2,4 Kepler GPUs. The six lines correspond to the CG algorithm: with preconditioning **i** (red and green lines), with preconditioning **iv** (blue and magenta lines) and with preconditioning **iii** in single precision (cyan and yellow lines).

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 Kepler K20) interconnected by InfiniBand.

Our results can be seen in the two plots in Fig. 1. 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, left plot) and as a function of the number of GPUs at a fixed lattice size (strong scaling, right plot).

Our data show a large overhead due to inter-GPU communication, especially for the last two preconditioners. This overhead is visibly reduced by using single precision.