

**Preconditioned Eigenvalue Solvers  
for electronic structure calculations**

**Andrew V. Knyazev**

*Department of Mathematics and  
Center for Computational Mathematics  
University of Colorado at Denver*

Householder Symposium XVI  
May 26, 2005

*Partially supported by the NSF and the LLNL*

## Abstract

We describe the locally optimal block preconditioned conjugate gradient (LOBPCG) method in the framework of ABINIT and Vienna Ab-initio Simulation Package (VASP). Several methods are available in ABINIT/VASP to calculate the electronic ground state: simple Davidson-block iteration scheme, single band steepest descent scheme, conjugate gradient optimization, residual minimization method. LOBPCG can be interpreted as a conjugate gradient optimization, different from those used in ABINIT/VASP. It can also be viewed as the steepest descent scheme, augmented with extra vectors in the basis set, namely with the wave functions from the previous iteration step, not with the residuals as implemented in VASP. Finally, it can be seen as simplified specially restarted block Davidson method. We describe the LOBPCG and compare it with ABINIT/VASP algorithms.

## Acknowledgements

Partially supported by the NSF and the LLNL Center for Applied Scientific Computing.

We would like to thank Rob Falgout, Charles Tong, Panayot Vassilevski and other members of the Hypre Scalable Linear Solvers LLNL project team for their support in implementing our LOBPCG eigensolver in Hypre.

John Pask and Jean-Luc Fattebert of LLNL provided invaluable help in getting us involved in the electronic structure calculations.

LOBPCG is implemented in ABINIT rev. 4.5 and above by G. Zérah.

## LOBPCG for electronic structure calculations

1. Electronic Structure Calculations
2. The basics of ABINIT and VASP
3. VASP band by band preconditioned steepest descent
4. ABINIT/VASP band by band preconditioned conjugate gradients
5. Locally optimal band by band preconditioned conjugate gradients
6. VASP simple block Davidson (block preconditioned steepest descent)
7. ABINIT LOBPCG
8. Conclusions

## Electronic Structure Calculations

Quantum mechanics: a system of interacting electrons and nuclei described by the many-body wavefunction  $\Psi$  (with  $M^N$  complexity, where  $M$  is the number of space points and  $N$  is the number of electrons), a solution to the Schrodinger equation  $H\Psi = E\Psi$ , where  $H$  is the Hamiltonian operator and  $E$  the total energy of the system. The Hamiltonian operator contains the kinetic operators of each individual electron and nuclei and all pair-wise Coulomb interactions. Density Functional Theory and Kohn and Sham: the many-body Schrodinger equation can be reduced to an effective one-particle equation:  $[-\nabla^2 + v_{eff}(\mathbf{r})] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r})$ , which depends on the electronic density  $\sum_{j=1}^N |\psi_j(r)|^2$  (self-consistent) with  $N$  occupied and mutually orthogonal single particle wave functions  $\psi_i$ ,  $M \times N$  total complexity.

<http://www.ABINIT.org/> “First-principles computation of material properties: the ABINIT software project,” X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanesse, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph. Ghosez, J.-Y. Raty, D.C. Allan. Computational Materials Science 25, 478-492 (2002)

<http://cms.mpi.univie.ac.at/vasp/> “Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set,” G. Kresse and J. Furthmuller, Comput. Mat. Sci. 6, 15-50 (1996)

<http://math.cudenver.edu/~aknyazev/> A. V. Knyazev, Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. SIAM J. Sci. Comput. 23, no. 2, 517–541 (2001)

## The basics of ABINIT and VASP

ABINIT and Vienna Ab-initio Simulation Package (VASP) are complex packages for performing ab-initio quantum-mechanical molecular dynamics simulations using a plane wave (PW) basis set. Generally not more than 100 PW per atom are required to describe bulk materials, in most cases even 50 PW per atom will be sufficient for a reliable description.

Multiplication by the Hamiltonian in the plain wave basis involves both Fourier (for the Laplacian) and real (for the potential) spaces, thus DFFT is called twice per multiplication. Since the Laplacian becomes diagonal in the Fourier space, diagonal preconditioning is useful.

Alternatively, real-space FEM and finite difference codes are available, e.g., Multigrid Instead of the K-space (MIKA). In real space approximations the number of basis functions is 10-100 times larger than that with plain waves, and sophisticated preconditioning (usually multigrid) is necessary.

The execution time scales like  $CMN^2$  for some parts of the code, where  $M$  is the number of PW basis functions and  $N$  is the number of valence electrons in the system. The constant  $C$  is controlled by [keeping the number of orthogonalizations small](#). For systems with roughly 2000 electronic bands, the  $CMN^2$  part in VASP becomes comparable to other parts. Systems with up to 4000 valence electrons can be solved.

VASP and ABINIT use a self-consistency cycle to calculate the electronic ground-state of the Kohn-Sham equation.

The Kohn-Sham (KS) eigenvalue problem (eigenproblem) to solve is

$$\mathbf{H}|\phi_n\rangle = \epsilon_n \mathbf{S}|\phi_n\rangle, n = 1, \dots, N_b,$$

where  $\mathbf{H}$  is the Kohn-Sham Hamiltonian,  $|\phi_n\rangle$  is the band eigenstate wave function,  $\epsilon_n$  is the eigenvalue, describing the band energy, and  $\mathbf{S}$  is the overlap matrix, specifying the orthonormality constraints of eigenstates for different bands. In PW basis, the calculations of  $\mathbf{H}|\phi_n\rangle$  and  $\mathbf{S}|\phi_n\rangle$  are relatively inexpensive. Matrices  $\mathbf{H}$  and  $\mathbf{S}$  are Hermitian and  $\mathbf{S}$  is positive definite. In VASP,  $\mathbf{S}$  is not identity because of the use of ultra-soft pseudo potentials.  $N_b \ll N$  wavefunctions of occupied orbitals are computed. In a self-consistent calculation, optimization of the charge density, appearing in the Hamiltonian, and wavefunctions  $|\phi_n\rangle$  of the KS eigenproblem is done in a cycle. In this talk, we consider only the KS eigenproblem numerical solution part, not the charge density updates.

The expectation value of the Hamiltonian for a wavefunction  $|\phi_{app}\rangle$

$$\epsilon_{app} = \frac{\langle \phi_{app} | \mathbf{H} | \phi_{app} \rangle}{\langle \phi_{app} | \mathbf{S} | \phi_{app} \rangle}$$

is called the Rayleigh quotient. Variation of the Rayleigh quotient with respect to  $\langle \phi_{app} |$  leads to the residual vector defined as

$$|R(|\phi_{app}\rangle)\rangle = (\mathbf{H} - \epsilon_{app}\mathbf{S})|\phi_{app}\rangle.$$

To accelerate the convergence, a preconditioner  $\mathbf{K}$  is applied to the residual:

$$|p(|\phi_{app}\rangle)\rangle = \mathbf{K}|R(|\phi_{app}\rangle)\rangle = \mathbf{K}(\mathbf{H} - \epsilon_{app}\mathbf{S})|\phi_{app}\rangle.$$

The preconditioner  $\mathbf{K}$  is Hermitian, often simply an inverse to the main diagonal of  $\mathbf{H} - \epsilon_{app}\mathbf{S}$  or of  $\mathbf{H}$  - the latter is used in our tests for this talk.

Another core technique used in VASP is the Rayleigh-Ritz variational scheme in a subspace spanned by  $N_a$  wavefunctions  $|\phi_i\rangle$ ,  $i = 1, \dots, N_a$ :

$$\langle \phi_j | \mathbf{H} | \phi_i \rangle = H_{ij}, \quad \langle \phi_j | \mathbf{S} | \phi_i \rangle = S_{ij}, \quad H_{ij} U_{jk} = \epsilon_k S_{ij} U_{jk}, \quad |\phi_j\rangle = U_{jk} |\phi_k\rangle,$$

using the standard summation agreement on repeated indexes from 1 to  $N_a$ . Here, on the first two steps we compute the  $N_a$ -by- $N_a$  Gram matrices of the KS quadratic forms with respect to the given basis of wavefunctions. Then we solve a matrix generalized eigenproblem with the Gram matrices to compute the Ritz values  $\epsilon_k$ . Finally, we calculate the Ritz functions  $|\phi_j\rangle$ .

The set of wavefunctions here is not assumed to be  $\mathbf{S}$ -orthonormal and may include the approximate KS eigenfunctions as well as preconditioned residuals and other functions.

## VASP single band preconditioned steepest descent

The method is *robust and simple*. It utilizes the *two-term recurrence*

$$|\phi^{(M+1)}\rangle \in \text{Span} \{ |p^{(M)}\rangle, |\phi^{(M)}\rangle \},$$

where  $|p^{(M)}\rangle = |p(|\phi^{(M)}\rangle)\rangle$  is the preconditioned residual on the  $M$ -th iteration and  $|\phi^{(M+1)}\rangle$  is chosen to minimize the Rayleigh quotient  $\epsilon^{(M+1)}$  by using the *Rayleigh–Ritz method* on this two dimensional subspace. It can only compute approximately the lowest eigenstate  $|\phi_1\rangle$ . The asymptotic convergence speed is determined, when the preconditioner  $\mathbf{K}$  is positive definite, by the ratio  $\kappa$  of the largest and the smallest positive eigenvalues of  $\mathbf{K}(\mathbf{H} - \epsilon_1\mathbf{S})$ .

Two-term recurrence + Rayleigh–Ritz method =  
Single Band Steepest Descent Method

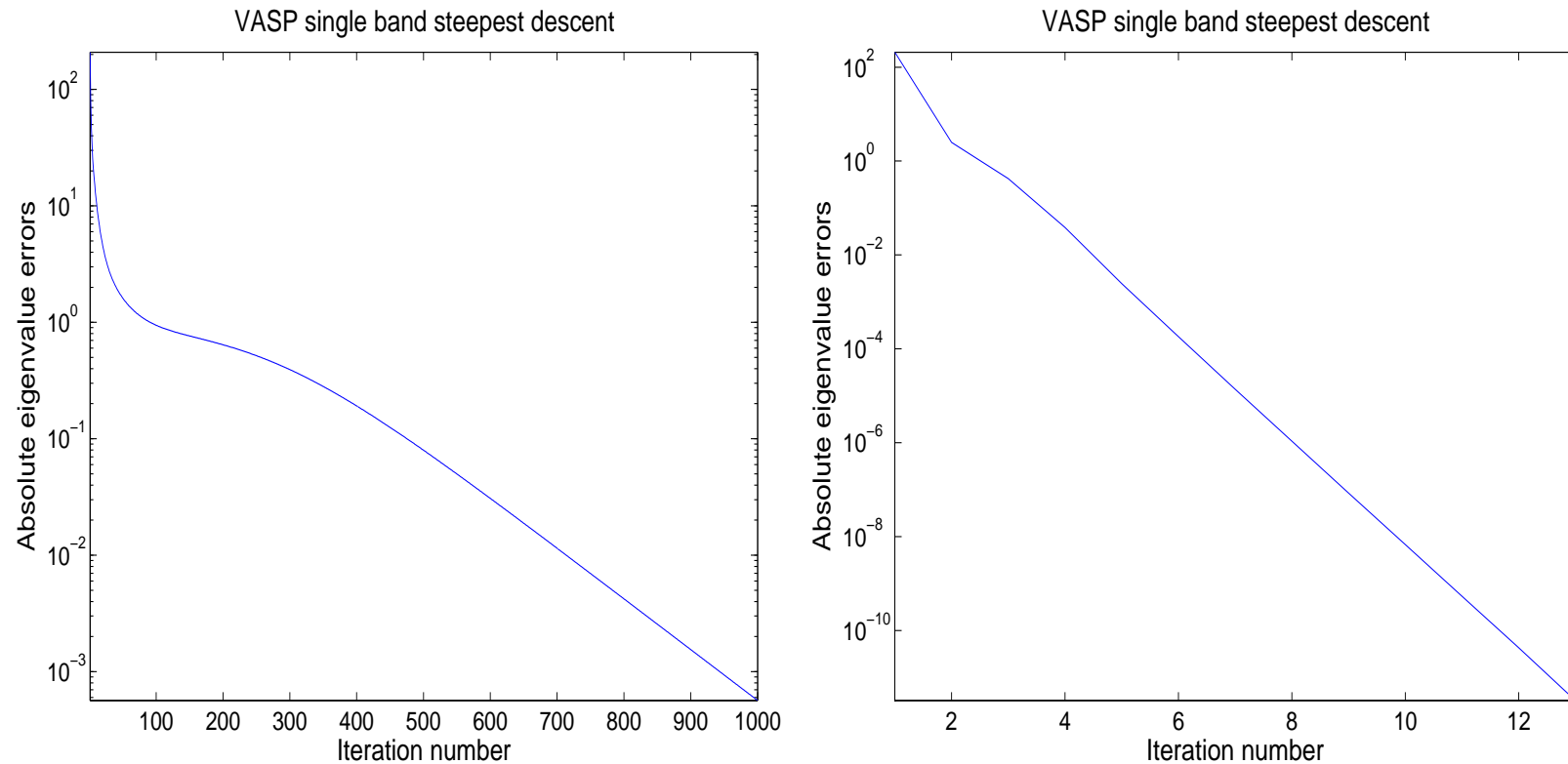


FIGURE 1: *SD*  $\kappa \approx 4 * 10^2$  (left) and with diagonal preconditioning  $\kappa \approx 2$  (right), no cluster of eigenvalues in these tests

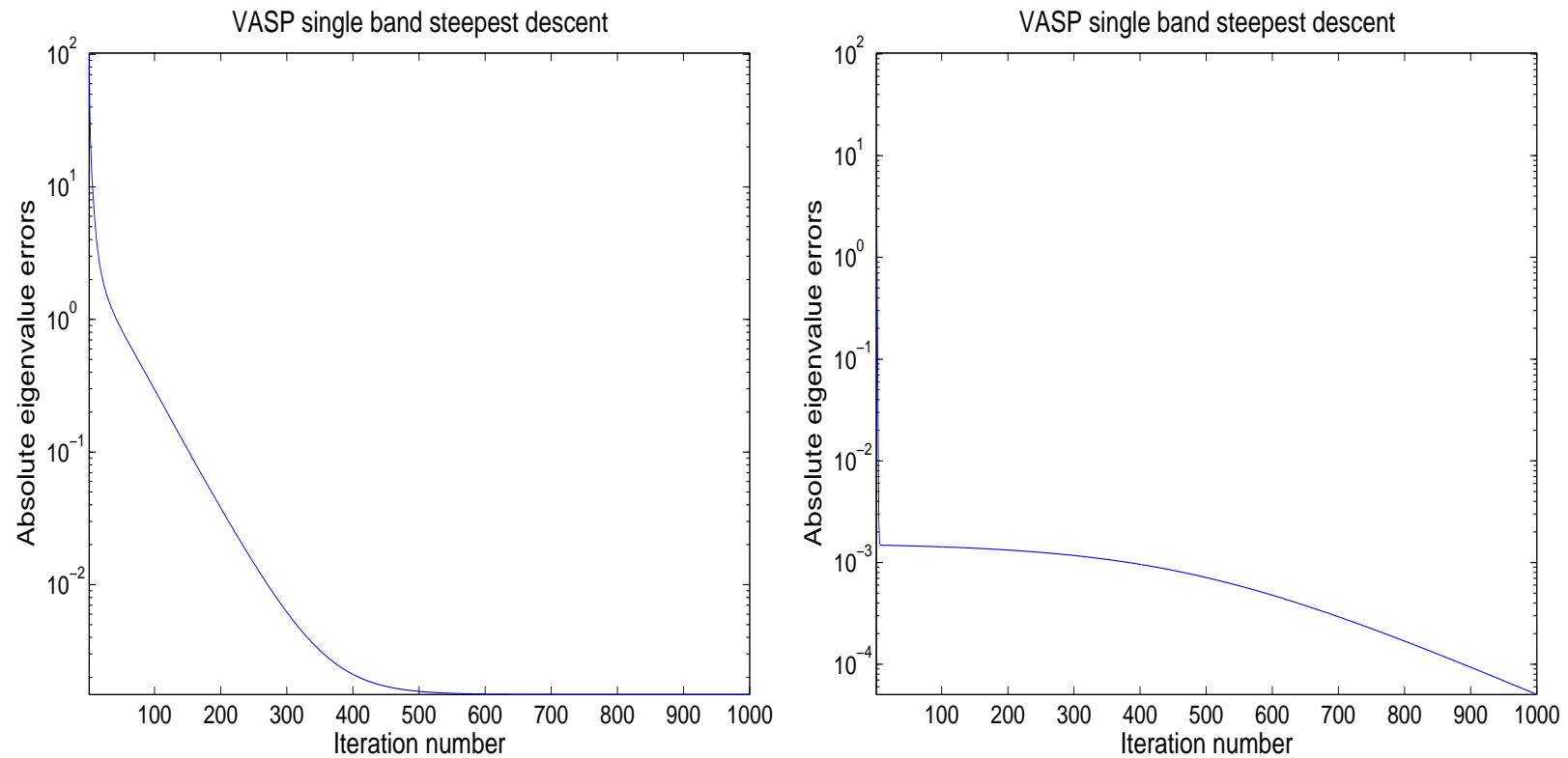


FIGURE 2: *SD*  $\kappa \approx 10^5$  (left) and with diagonal preconditioning  $\kappa \approx 6 \times 10^2$  (right) for an eigenvalue from a cluster of  $\approx 10^{-3}$  width. The fast convergence in the beginning until the cluster resolution.

## VASP band by band preconditioned steepest descent

After we computed the approximate lowest eigenstate  $|\phi_1\rangle$ , e.g., using the VASP single band preconditioned steepest descent method from the previous slide, we can compute the next lowest by taking the  $\mathbf{S}$ -orthogonal complement to  $|\phi_1\rangle$ . We need to orthogonalize the initial guess, and the current preconditioned residual vector  $|p^{(M)}\rangle$  is being replaced with  $|g^{(M)}\rangle = (1 - |\phi_1\rangle\langle\phi_1|\mathbf{S})|p^{(i)}\rangle$ . If we use again the single band preconditioned steepest descent in the  $\mathbf{S}$ -orthogonal complement, we can compute a number of lowest eigenstates band by band. Asymptotic convergence is **not cluster robust**. **No guarantee for cluster calculations!**

Singe Band Preconditioned Method +

Orthogonalization to the Previously Computed Eigenstates =

Band by Band Preconditioned Method

## Single band preconditioned conjugate gradients

A well known approach to accelerate convergence of the preconditioned steepest descent is to replace it with the preconditioned conjugate gradients (PCG). The Rayleigh quotient that we want to minimize is a smooth function for nonzero wavefunctions and its gradient and its Hessian are easy to calculate explicitly. Any version of the PCG for smooth non-quadratic functions can be tried for the minimization of the Rayleigh quotient. VASP apparently (the method description is too sketchy) uses a version known in optimization as the Fletcher–Reeves formula:  $|f^{(0)}\rangle = 0$ ,

$$|f^{(M)}\rangle = |p^{(M)}\rangle + \text{sign}(\gamma_\phi^{(M)}) \frac{\langle p^{(M)} | R^{(M)} \rangle}{\langle p^{(M-1)} | R^{(M-1)} \rangle} |f^{(M-1)}\rangle.$$

$f^{(M)}$  replaces  $p^{(M)}$  in the two dimensional trial subspace so that  $|\phi^{(M+1)}\rangle = \gamma_f^{(M)} |f^{(M)}\rangle + \gamma_\phi^{(M)} |\phi^{(M)}\rangle$  in the Rayleigh–Ritz procedure.

## Band by band preconditioned conjugate gradients

After we computed the approximate lowest eigenstate  $|\phi_1\rangle$ , e.g., using the single band preconditioned conjugate gradients method from the previous slide, we can compute the next lowest by taking the  $\mathbf{S}$ -orthogonal complement to  $|\phi_1\rangle$ . We need to orthogonalize the initial guess, and the current preconditioned residual vector  $|p^{(M)}\rangle$  is being replaced with  $|g^{(M)}\rangle = (1 - |\phi_1\rangle\langle\phi_1|\mathbf{S})|p^{(i)}\rangle$ . Fletcher–Reeves formula changes to:

$$|f^{(M)}\rangle = |g^{(M)}\rangle + \text{sign}(\gamma_\phi^{(M)}) \frac{\langle g^{(M)} | R^{(M)} \rangle}{\langle g^{(M-1)} | R^{(M-1)} \rangle} |f^{(M-1)}\rangle, |f^{(0)}\rangle = 0.$$

and  $|\phi^{(M+1)}\rangle = \gamma_f^{(M)} |f^{(M)}\rangle + \gamma_\phi^{(M)} |\phi^{(M)}\rangle$  in the Rayleigh–Ritz procedure.

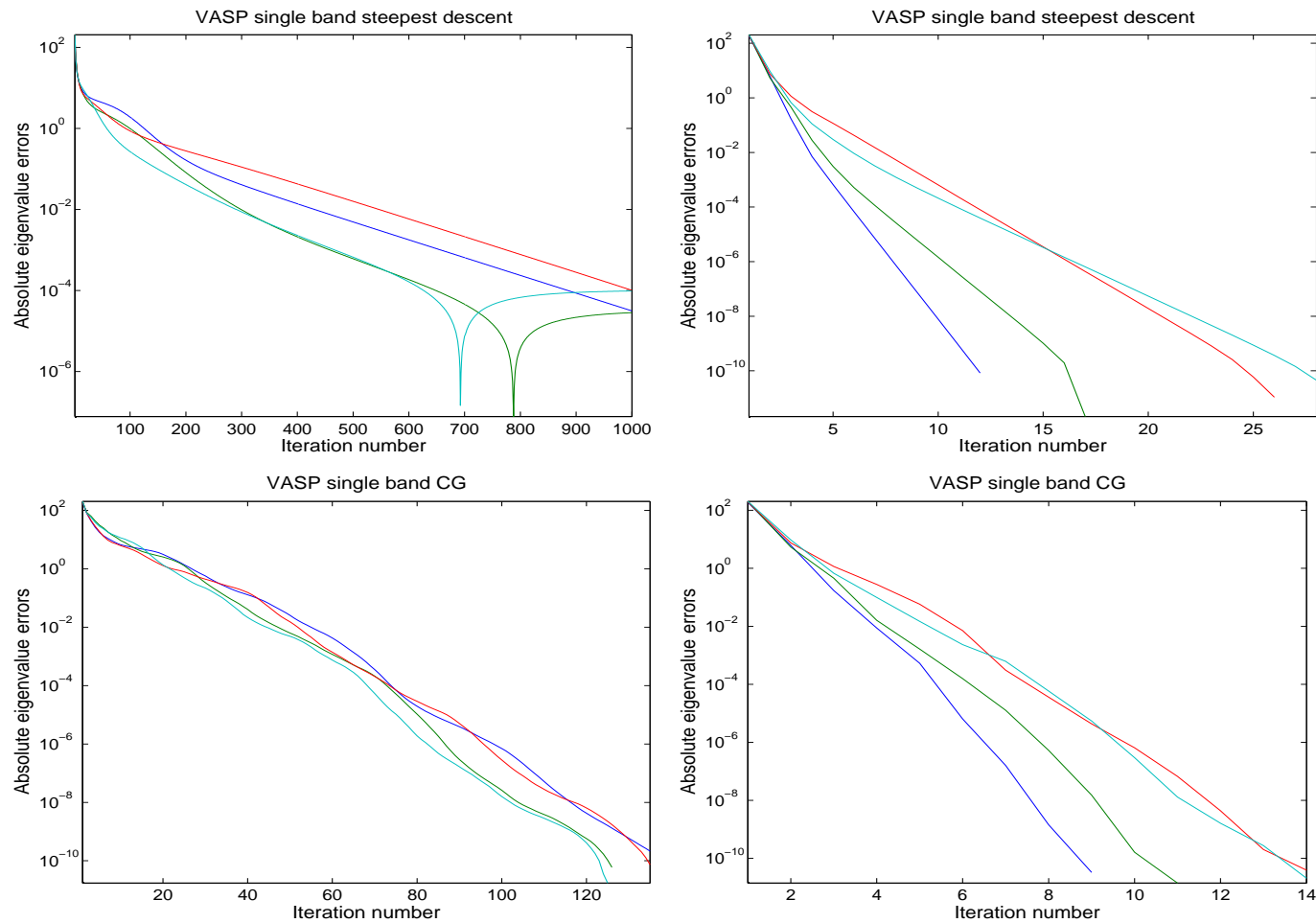


FIGURE 3: *SD and CG without (left) and with (right) preconditioning.*

## Locally optimal single band PCG

The method combines *robustness and simplicity* of the preconditioned steepest descent method with a *three-term recurrence* formula:

$$|\phi^{(M+1)}\rangle \in \text{Span} \{ |p^{(M)}\rangle, |\phi^{(M)}\rangle, |\phi^{(M-1)}\rangle \},$$

where  $|p^{(M)}\rangle = |p(|\phi^{(M)}\rangle)\rangle$  is the preconditioned residual on the  $M$ -th iteration and  $|\phi^{(M+1)}\rangle$  is chosen to minimize the Rayleigh quotient  $\epsilon^{(M+1)}$  by using the *Rayleigh–Ritz method* on this three dimensional subspace. The asymptotic convergence speed to lowest eigenstate  $|\phi_1\rangle$  depends, when the preconditioner  $\mathbf{K}$  is positive definite, on the *square root* of the ratio  $\kappa$  of the largest and the smallest positive eigenvalues of  $\mathbf{K}(\mathbf{H} - \epsilon_1\mathbf{S})$ .

Three-term recurrence + Rayleigh–Ritz method =  
Locally Optimal Conjugate Gradient Method

## Locally optimal band by band PCG

After we computed the approximate lowest eigenstate  $|\phi_1\rangle$ , e.g., using the locally optimal single band preconditioned conjugate gradients method from the previous slide, we can compute the next lowest by taking the  $\mathbf{S}$ -orthogonal complement to  $|\phi_1\rangle$ . We need to orthogonalize the initial guess, and the current preconditioned residual vector  $|p^{(M)}\rangle$  is being replaced with  $|g^{(M)}\rangle = (1 - |\phi_1\rangle\langle\phi_1|\mathbf{S})|p^{(i)}\rangle$ . No other changes are needed.

The band by band implementation of the locally optimal PCG method shares the same advantages and difficulties with other band by band iterative schemes, such as low memory requirements and troubles resolving clusters of eigenvalues.

## Locally optimal vs. Fletcher–Reeves PCG

The methods converge quite similarly in practice in many cases, though a theoretical explanation of this fact is still not available. The costs per iterations are also similar, but the locally optimal method is slightly more expensive. There is known (Knyazev, 2001) a reasonably stable implementation of the locally optimal method that uses an implicitly computed linear combination of  $|\phi^{(M)}\rangle$ ,  $|\phi^{(M-1)}\rangle$  instead  $|\phi^{(M-1)}\rangle$  in the Rayleigh-Ritz procedure. The convergence of the locally optimal method is well supported theoretically, while the convergence of the Fletcher–Reeves version is not.

Perhaps, the main practical advantage of the locally optimal method is that it allows a straightforward block generalization to calculate several eigenpairs simultaneously.

## Block eigenvalue iterations

A well known idea of using Simultaneous, or Block Iterations provides an important improvement over single-vector methods, and permits us to compute an invariant subspace, rather than one eigenvector at a time, which allows robust calculations of clustered eigenvalues and corresponding invariant subspaces. It can also serve as an acceleration technique over a single-vector methods on parallel computers, as convergence for extreme eigenvalues usually increases with the size of the block, and every step can be naturally implemented on wide varieties of multiprocessor computers as well as to take advantage processors cache through the use of high level BLAS libraries.

## VASP simple block Davidson (Block preconditioned steepest descent)

$$|\phi_i^{(M+1)}\rangle \in \text{Span} \{ |p_1^{(M)}\rangle, \dots, |p_{N_b}^{(M)}\rangle, |\phi_1^{(M)}\rangle, \dots, |\phi_{N_b}^{(M)}\rangle \}, i = 1, \dots, N_b,$$

where  $|p_i^{(M)}\rangle = |p(|\phi_i^{(M)}\rangle)\rangle$  is the preconditioned residual on the  $M$ -th iteration for the  $i$ th band and  $|\phi_i^{(M+1)}\rangle$  are chosen as Ritz functions by using the *Rayleigh–Ritz method* on this  $2N_b$  dimensional subspace, to approximate the  $N_b$  lowest exact KS eigenstates  $|\phi_i\rangle$ . The asymptotic convergence speed for the  $i$ th band is determined, when the preconditioner  $\mathbf{K}$  is positive definite, by the ratio  $\kappa_i$  of the largest and the smallest positive eigenvalues of  $\mathbf{K}(\mathbf{H} - \epsilon_i \mathbf{S})$ .

Trial subspace of all approx. eigenfunctions and residuals +

Rayleigh–Ritz method = Block Steepest Descent Method

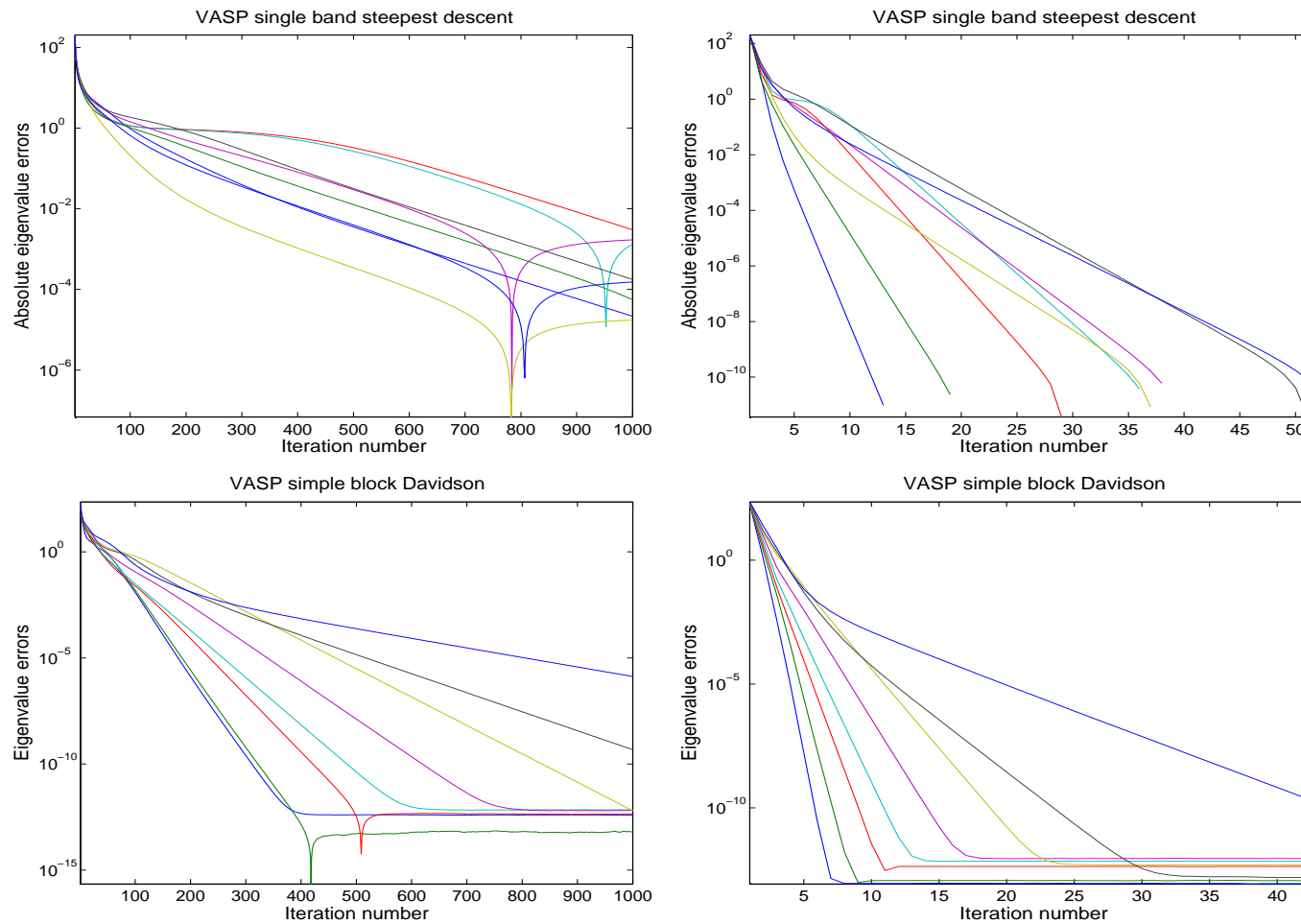


FIGURE 4: *SD and block SD without (left) and with (right) preconditioning.*

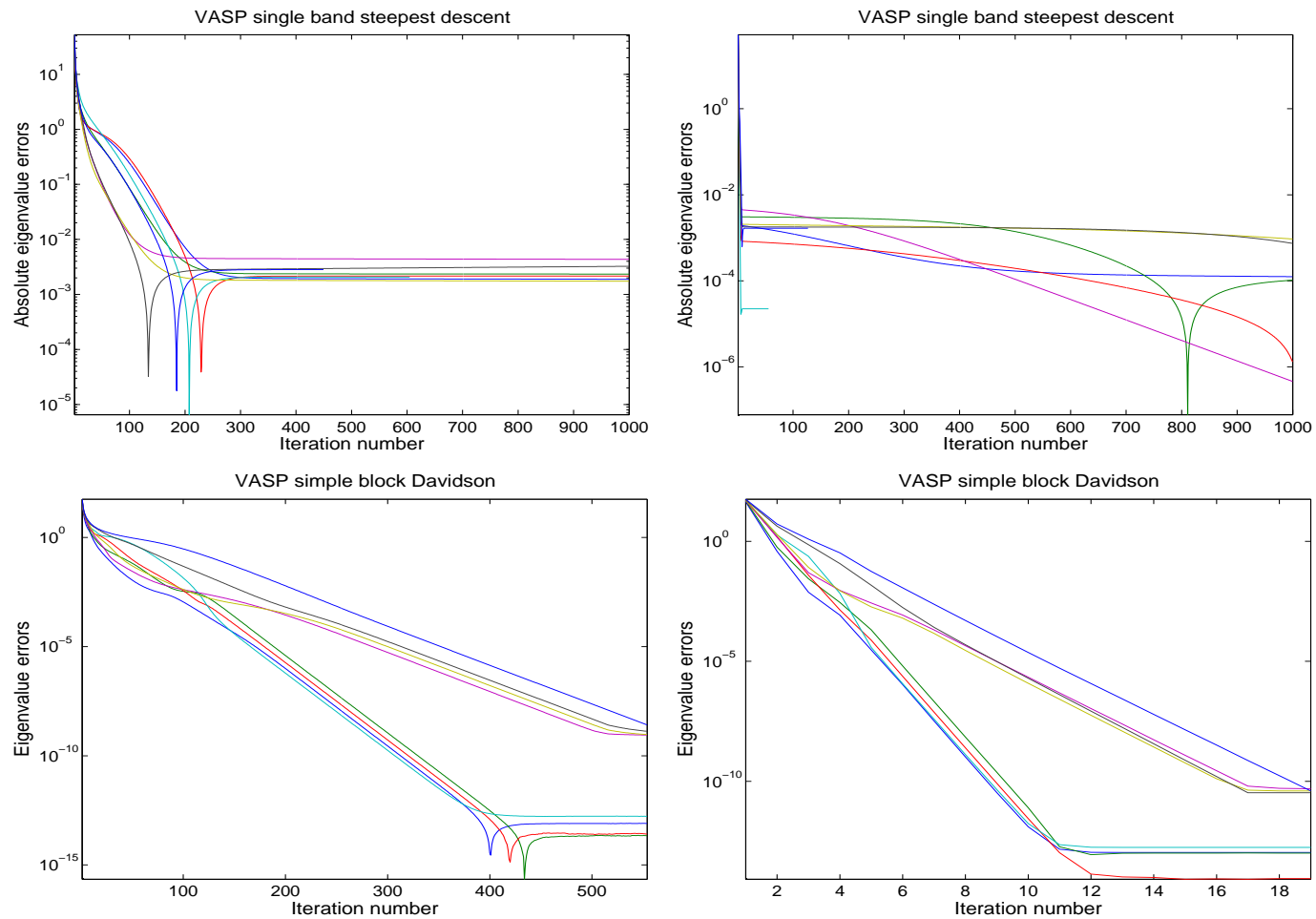


FIGURE 5: *SD and block SD without (left) and with (right) preconditioning.*

## General block Davidson with a trivial restart

The block preconditioned steepest descent is the simplest case of a general block Davidson method. In the general block Davidson method, the trial subspace of the Rayleigh–Ritz method increases in each step by  $N_b$  current preconditioned residuals, i.e., in its original formulation the general block Davidson method memory requirements grow with every iteration by  $N_b$  vectors. Such method converges faster than the block steepest descent, but the costs are growing with the number of iterations.

A restart is needed in practice. A trivial restart at every step restricts the trial subspace only to the  $N_b$  approximate eigenfunctions and one set of  $N_b$  preconditioned residuals, thus, leading to the block preconditioned steepest descent from the previous slide, utilized in VASP.

## General block Davidson with a non-standard restart

A non-standard restart in the block Davidson method is suggested in C. Murray, S. C. Racine, E. R. Davidson, Improved algorithms for the lowest few eigenvalues and associated eigenvectors of large matrices. *J. Comput. Phys.* 103 (1992), no. 2, 382–389.

A. V. Knyazev, Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. *SIAM J. Sci. Comput.* 23 (2001), no. 2, 517–541 (electronic).

In addition to the usual  $N_b$  current approximate eigenfunctions  $|\phi_1^{(M)}\rangle, \dots, |\phi_{N_b}^{(M)}\rangle$ , we also add  $N_b$  approximate eigenfunctions from the previous step:  $|\phi_1^{(M-1)}\rangle, \dots, |\phi_{N_b}^{(M-1)}\rangle$ .

## Locally optimal block PCG (LOBPCG)

Approximate eigenfunctions  $|\phi_i^{(M+1)}\rangle$  are chosen from the

$$\text{Span} \{ |p_1^{(M)}\rangle, \dots, |p_{N_b}^{(M)}\rangle, |\phi_1^{(M)}\rangle, \dots, |\phi_{N_b}^{(M)}\rangle, |\phi_1^{(M-1)}\rangle, \dots, |\phi_{N_b}^{(M-1)}\rangle \},$$

where  $|p_i^{(M)}\rangle = |p(|\phi^{(M)}\rangle)\rangle$  is the preconditioned residual on the  $M$ -th iteration for the  $i$ th band and  $|\phi_i^{(M+1)}\rangle$  are chosen as Ritz functions by using the *Rayleigh–Ritz method* on this  $3N_b$  dimensional subspace, to approximate the  $N_b$  lowest exact KS eigenstates  $|\phi_i\rangle$ . The asymptotic convergence speed for the  $i$ th band depends, when the preconditioner  $\mathbf{K}$  is positive definite, on the square root of the ratio  $\kappa_i$  of the largest and the smallest positive eigenvalues of  $\mathbf{K}(\mathbf{H} - \epsilon_i\mathbf{S})$ .

Current and one set of previous approx. eigenfunctions plus residuals +

Rayleigh–Ritz method = Block Conjugate Gradient Method

Center for Computational Mathematics, University of Colorado at Denver

## LOBPCG interpretations:

- LOBPCG can be interpreted as a specific version of conjugate gradient optimization, different from those used in VASP, that can be easily applied in the block form.
- It can also be viewed as the block steepest descent scheme, augmented with extra vectors in the basis set, namely with the wavefunctions from the previous iteration step, not with the residuals as in the block Davidson method.
- Finally, it can be seen as simplified specially restarted block Davidson method. It seems to preserve the fast convergence of the block Davidson at the much lower costs.

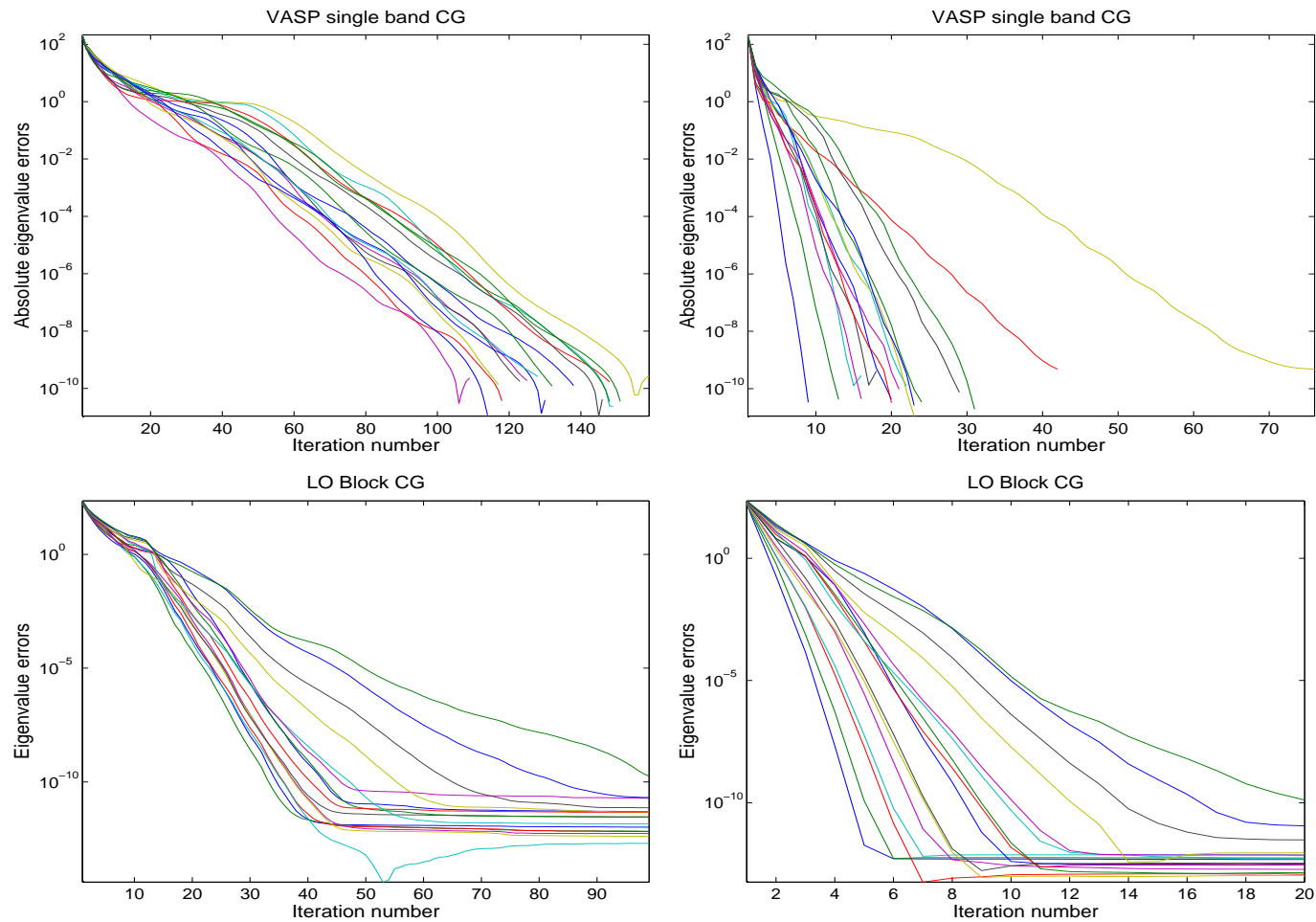


FIGURE 6: *CG and block CG without (left) and with (right) preconditioning.*

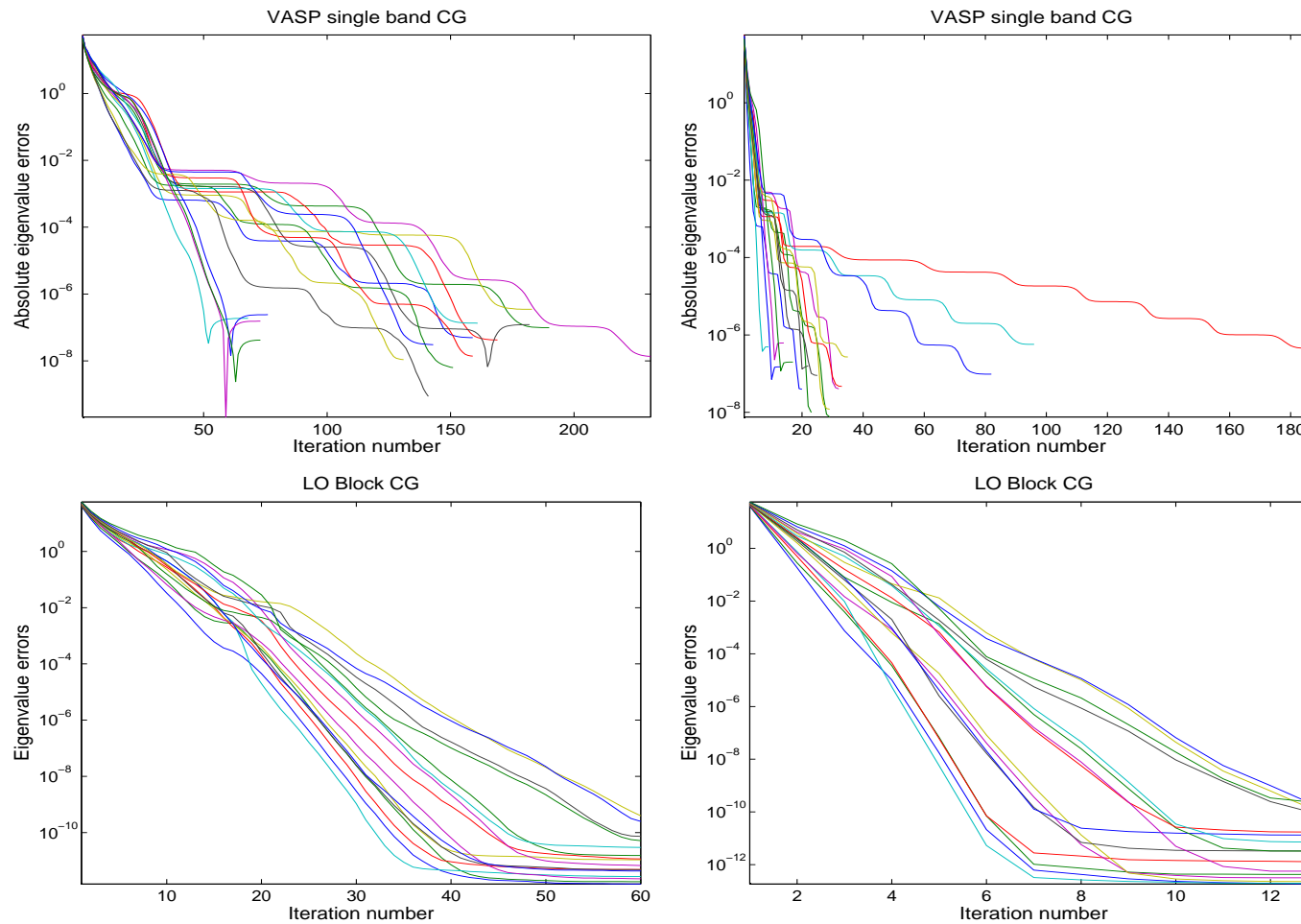


FIGURE 7: *CG and block CG without (left) and with (right) preconditioning.*

## Implementing LOBPCG in ABINIT

ABINIT is a FORTRAN 90 parallel code with timing. ABINIT rev. 4.5 manual: “The algorithm LOBPCG is available as an alternative to band-by-band conjugate gradient, from Gilles Zérah. Use `wfoptalg=4` . See Test v4 No. 93 and 94. The band-by-band parallelization of this algorithm should be much better than the one of the usual algorithm.” The implementation follows my MATLAB code of LOBPCG rev. 4.10 with some simplifications.

## Testing LOBPCG in ABINIT

Test example results provided by G. Zérah: 32 boron atoms and 88 bands, nonselfconsistent. Band-by-band PCG every band converges in 4-6 iterations. LOBPCG block size 88 converges in 2-4 iterations, approximately 2 times faster. LOBPCG block size 20 with partial convergence converges (we do not wait for the previous block to converge to compute the next, by limiting the number of LOBPCG iteration to  $nline=4$ ) is surprisingly even faster. We can even limit  $nline=2$  and get the same overall complexity as for  $nline=4$ .

Mixed results for LOBPCG II applied to the selfconsistent calculation of Hydrogen and nonselfconsistent 32 boron atoms: slow convergence for some bands.

## LOBPCG in PETSCAN

Work by others on LOBPCG in material sciences:

“Comparison of Nonlinear Conjugate-Gradient Methods for Computing the Electronic Properties of Nanostructure Architectures,” Stanimire Tomov, Julien Langou, Andrew Canning, Lin-Wang Wang, and Jack Dongarra implement and test LOBPCG in PESCAN, where a semi-empirical potential or a charge patching method is used to construct the potential and only the eigenstates of interest around a given energy are calculated using folded spectrum: “if memory is not a problem and block version of the matrix-vector multiplication can be efficiently implemented, the FS-LOBPCG will be the method of choice for the type of problems discussed.”

## Conclusions

- LOBPCG is a valuable alternative to eigensolvers currently used in electronic structure calculations
- Initial numerical results look promising, but more testing is needed on larger problems

## Future work

- ABINIT LOBPCG testing vs. the default CG solver
- New LOBPCG modifications with reduced orthogonalization costs
- Revisiting traditional preconditioning
- Collaboration with John Pask of LLNL on implementing LOBPCG in the in-house FEM electronic structure calculations code