Numerical methods

- Solving the Kohn-Sham Equations
- SCF
- Geometry Optimization

Shobhana Narasimhan

JNCASR, Bangalore, India

UCSB-ICMR Summer School on Electronic Structure Calculations

Santa Barbara, August 2005

The Kohn-Sham problem

- Minimize energy functional:

$$E[\{\psi_i\}] = \sum_i \int \psi_i [\frac{1}{2}] \nabla^2 \psi_i d^3r + \int V_{ion} n(\mathbf{r}) d^3r +$$

$$\frac{1}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^3r d^3r' + E_{XC}[n(\mathbf{r})]$$

$$+ V_{ion}(\{\mathbf{R}_I\})$$

- Or solve Kohn-Sham equations:

$$\left[\tfrac{1}{2}\nabla^2 + V_{ion}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r})$$

- Note that self-consistent solution necessary, as H depends on solution: $\{\psi_i\} \rightarrow n(\mathbf{r}) \rightarrow H$

- Convention: $e = \hbar = m_e = 1$

Numerical Solutions:
Factors to keep in mind

- Want: Accurate but cheap…

- Memory requirements should be small.

- CPU time should be short.

  Iterative solutions:

  - Number of iterations should be small (fast convergence).

  - Each iteration should be fast.

- For parallel architectures: good scalability.

- Recall (from yesterday): pseudopotentials were introduced to speed the calculation, but need to do more!

Warnings / Disclaimers

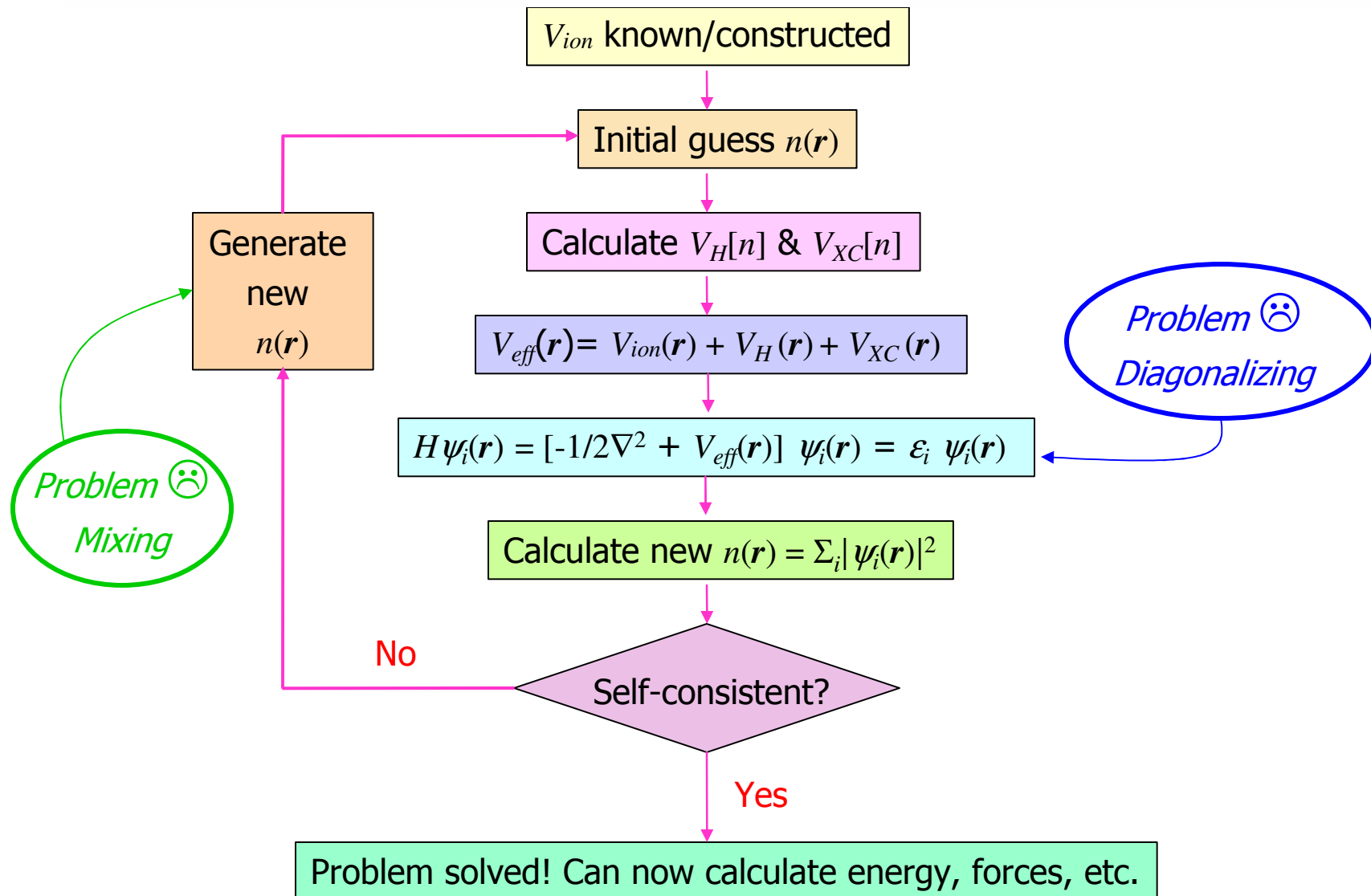For simplicity / ease of presentation, will make following assumptions:

- Try to focus on general concepts rather than mathematical formulae.

- Not exhaustive, there exist many schemes!

- Wavefunctions expanded in a plane wave basis of size N_{PW}:

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{q}} c_{i,\mathbf{q}} \times \frac{1}{\sqrt{\Omega}}\exp(i\mathbf{q} \cdot \mathbf{r}).$$

  (Umesh Waghmare will discuss.)

- Assume $V_{ext} = V_{ion} = V_{PP}$

- Drop spin indices (will be present in spin-polarized calculations).

- Drop occupation factors (will be present for metals where there are partially occupied states).

Self-consistent solution: one way

V_{ion} known/constructed

↓

Initial guess $n(\mathbf{r})$

↓

Calculate $V_H[n]$ & $V_{XC}[n]$

↓

$V_{eff}(\mathbf{r})= V_{ion}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})$

↓

$H\psi_i(\mathbf{r}) = [-1/2\nabla^2 + V_{eff}(\mathbf{r})]\ \psi_i(\mathbf{r}) = \varepsilon_i\ \psi_i(\mathbf{r})$

↓

Calculate new $n(\mathbf{r}) = \Sigma_i|\psi_i(\mathbf{r})|^2$

↓

Self-consistent?

Generate new $n(\mathbf{r})$

No

Yes

Problem ☹ *Diagonalizing*

Problem ☹ *Mixing*

Problem solved! Can now calculate energy, forces, etc.

Hamiltonian in plane wave basis

- Expand wavefunction in plane wave basis:

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}} \times \frac{1}{\sqrt{\Omega}}\exp[i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}].$$

- Eigenvalue equation is now:

$$\sum_{\mathbf{G}'} H_{\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}'} c_{i,\mathbf{k}+\mathbf{G}'} = \epsilon_i c_{i,\mathbf{k}+\mathbf{G}}$$

- Matrix elements are:

$$\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \delta_{\mathbf{G},\mathbf{G}'} + \mathbf{V}_{\text{ion}}(\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}') + \mathbf{V}_{\text{H}}(\mathbf{G}-\mathbf{G}') + \mathbf{V}_{\text{XC}}(\mathbf{G}-\mathbf{G}')$$

- Note that in reciprocal space, kinetic energy term is diagonal.
- Ionic potential given by:

$$V_{ion}(\mathbf{G}) = \sum_{\alpha} \mathbf{S}_{\alpha}(\mathbf{G})\mathbf{v}_{\alpha}(\mathbf{G}); \quad \mathbf{S}_{\alpha}(\mathbf{G}) = \sum_{\text{I}}\exp(\mathrm{i}\mathbf{G}\cdot\mathbf{R}_{\text{I}})$$

Exact diagonalization is expensive!

- Expand in a basis of plane waves.

- Have to diagonalize (find eigenvalues & eigenfunctions of) $H_{\mathbf{k+G,k+G'}}$

- Typically, $N_{PW} > 100$ x number of atoms in unit cell.

- Expensive to store H matrix: N_{PW}^2 elements to be stored.

- Expensive (CPU time) to diagonalize matrix exactly, $\sim N_{PW}^3$ operations required.

- Note, $N_{PW} >> N_b$ = number of bands required = $N_e/2$ or a little more (for metals).
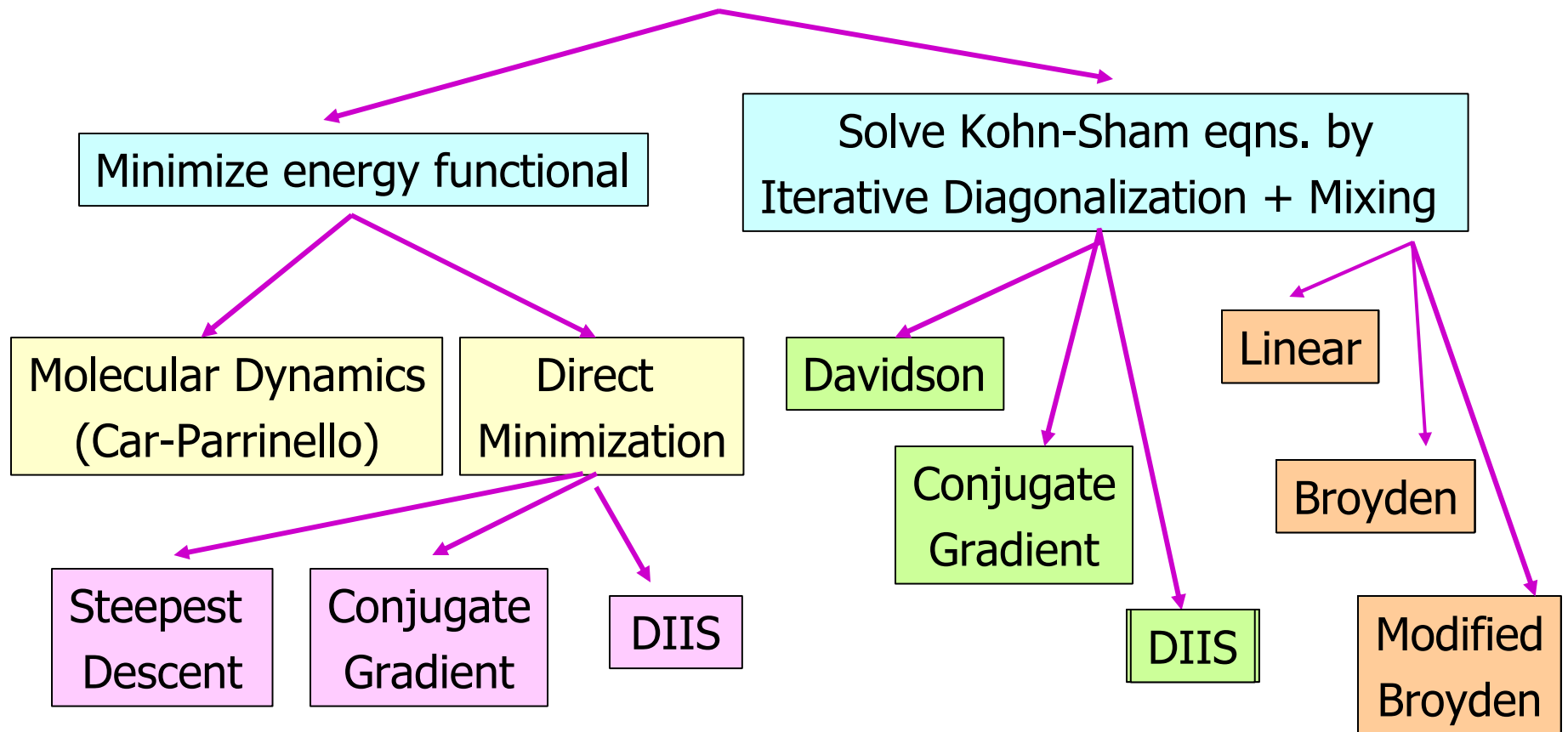
- So ok to determine just lowest few eigenvalues.

Initial Choice of n(r)

Various possible choices, e.g.,:

- Converged $n(\mathbf{r})$ from a closely related calculation
  (e.g., one where ionic positions slightly different).

- Superpositions of atomic densities.

- Approximate $n(\mathbf{r})$, e.g., from solving problem in a smaller/different basis.

- Random numbers.

Minimization

- Relevant to many parts of calculation:
  - Minimizing energy functional
  - Diagonalizing Hamiltonian
  - Achieving self-consistency
  - Optimizing ionic positions

- Consider a function $F(\mathbf{x})$ of N variables $\mathbf{x} = x_1, x_2, \ldots, x_N$

- Assume the gradient $\nabla F(\mathbf{x})$ is also known.
  $\nabla F(\mathbf{x})$ points in direction of maximum increase of $F(\mathbf{x})$.

- Want to find $\mathbf{x}_0$ s.t. $F(\mathbf{x})$ has its minimum value at $\mathbf{x}_0$, i.e., $\nabla F(\mathbf{x}_0)=0$.

- Will find iteratively, through a sequence of points $\mathbf{x}^n$ in the N-dimensional space that are stepping stones to finding $\mathbf{x}_0$.

- Convention: subscripts $\leftrightarrow$ coordinates, superscripts $\leftrightarrow$ iterations.

Steepest Descent

● Keep going downhill in direction opposite local gradient.

● Could try taking lots of little downhill steps:

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \alpha \, \nabla F(\mathbf{x}^n) = \mathbf{x}^n + \alpha \mathbf{g}^n$$

  - How to choose α?

    (too small: slow convergence, too large: overshoot)

● Always, $\mathbf{g}^n$ perpendicular to $\mathbf{g}^{n+1}$.

● Better: Once direction $\mathbf{g}^n$ identified, do line minimization (search along $\mathbf{g}^n$ for minimum or search for place where gradient is perpendicular to line), i.e. find optimal α^n.

● Convergence can be slow
(may not reach minimum!)

● Problem: when moving along new direction, lose some minimization along old one(s).
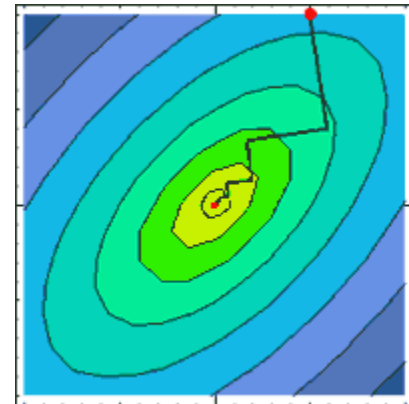
Quadratic Forms

- Consider a function of the form:

$$F(\{x_i\}) = F(\mathbf{x}) = \tfrac{1}{2}\mathbf{x} \cdot \mathbf{H} \cdot \mathbf{x}$$

- The Hessian is:

$$H_{ij} = \partial^2 F(\mathbf{x})/\partial x_i \partial x_j$$ (gives curvature)

- The (negative of) Gradient is then

$$\mathbf{g} = -\nabla F(x) = -\mathbf{H} \cdot \mathbf{x}$$

- Can approximate functions by quadratic form near minimum.

Note: for notational convenience, have dropped transpose signs...

$\mathbf{x}$ is an (Nx1) column vector, $\mathbf{H}$ is an (NxN) matrix, $F(\mathbf{x})$ is a scalar.

Quadratic Forms & Linear Equations

- Useful property:

  Minimizing the quadratic form:

  $$\tfrac{1}{2}\mathbf{x} \cdot \mathbf{H} \cdot \mathbf{x} - \mathbf{b} \cdot \mathbf{x} + c$$

  is equivalent to solving the linear system of equations:

  $$\mathbf{H} \cdot \mathbf{x} = \mathbf{b}$$

- Define: Error $\mathbf{e}^n = \mathbf{x}^n - \mathbf{x_o}$
- Define: Residual $\mathbf{r}^n = \mathbf{H} \cdot \mathbf{x}^n - \mathbf{H} \cdot \mathbf{x}_0$
- Algorithms can aim to minimize (norm of) error / residual

Conjugate Directions

● Want to find non-interfering directions s.t. minimization along one not spoiled by subsequent minimization along another.

● Conjugate directions are H-orthogonal:

$$\mathbf{d}^n \cdot H \cdot \mathbf{d}^m = 0, \quad \text{for} \quad n \neq m$$

● Can avoid having to know/store Hessian, instead
  - get new conjugate direction by:

$$\mathbf{d}^{n+1} = \mathbf{g}^{n+1} + \gamma^{n+1}\mathbf{d}^n$$

where

$$\gamma^{n+1} = \frac{\mathbf{g}^{n+1} \cdot \mathbf{g}^{n+1}}{\mathbf{g}^n \cdot \mathbf{g}^n} \quad \text{or} \quad \frac{(\mathbf{g}^{n+1} - \mathbf{g}^n) \cdot \mathbf{g}^{n+1}}{\mathbf{g}^n \cdot \mathbf{g}^n}$$

Fletcher-Reeves

Polak-Ribière

Conjugate Gradient Minimization

- Very first step: do steepest descent along gradient.
- Find conjugate direction.
- Do line minimization along conjugate direction.
- Find new conjugate direction; repeat until converged.
- If $F(\mathbf{x})$ can be written as a quadratic form, each new direction is conjugate to ALL previous directions, and minimum in N-d space found in at most N steps.
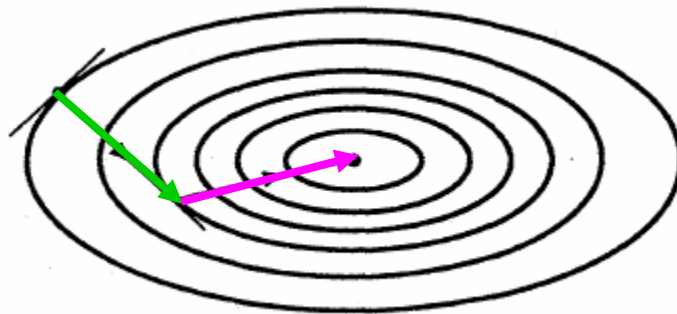
Preconditioning

- Preconditioning = extra step in a numerical method, designed to accelerate convergence.

- Let's try to motivate geometrically in this context…

- Curvature given by Hessian.

- Convergence fastest if curvature ~ uniform in all directions.
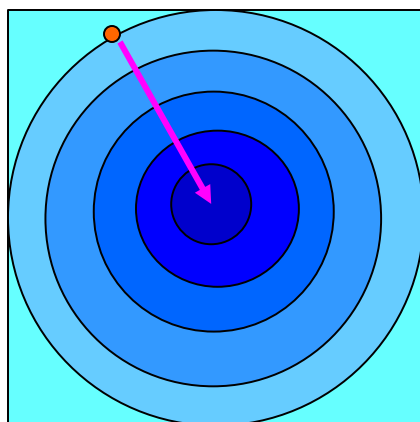
- For completely spherical function, converge in just 1 step.
- Want Hessian to look close to diagonal matrix with all elements same.
- Matrix eqs.: all eigenvalues clustered.

- Let $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x}$; so now minimize $\frac{1}{2}\,\mathbf{y}\cdot(\mathbf{PHP})\cdot\mathbf{y}$

- $\mathbf{P}^{-1}$ and multiplications easy (trivial) if $\mathbf{P}$ diagonal.

DIIS

- "Direct inversion of iterative subspace".
- Find a new trial vector $\mathbf{x}^{m+1}$ from a linear combination of all previous trial vectors $\mathbf{x}^1,\ldots,\mathbf{x}^m$:

$$\mathbf{x}^{m+1} = \sum_{i=1}^{m} c^i \mathbf{x}^i$$

- Ideally, new solution would be the right minimum:

$$\sum_{i=1}^{m} c^i \mathbf{x}^i = \mathbf{x}_0$$

- Define error $\mathbf{e}^i = \mathbf{x}^i - \mathbf{x}_0$. (In practice, would have to estimate).

DIIS [contd.]

- Substituting, we would like to have:

$$\sum_{i=1}^{m} c^i(x_0 + e^i) = \sum_{i=1}^{m} c^i x_0 + \sum_{i=1}^{m} c^i e^i = x_0$$

- Can enforce this if we set:

$$\sum_{i=1}^{m} c^i = 1 \qquad \text{and} \qquad \sum_{i=1}^{m} c^i e^i = 0$$

- Done by minimizing $\langle \sum_{i=1}^{m} c^i e^i | \sum_{j=1}^{m} c^j e^j \rangle$ subject to 1st constraint.

- Involves solving a $(m+1)$x$(m+1)$ matrix equation.
- Need some error estimate, e.g. (preconditioned) gradient...

Quasi-Newton-Raphson methods

- Want to minimize $F(\{x_i\}) = F(\mathbf{x}) = \frac{1}{2}\mathbf{x} \cdot \mathbf{H} \cdot \mathbf{x}$

- Gradients given by $\mathbf{g} = -\nabla F(x) = -\mathbf{H} \cdot \mathbf{x}$

- So $0 = \nabla F(\mathbf{x}_0) = \nabla F(\mathbf{x}^i) + \mathbf{H} \cdot (\mathbf{x}_0 - \mathbf{x}^i)$

- Can guess that minimum will be at:
$$\mathbf{x}^{i+1} = \mathbf{x}^i + \mathbf{H}^{-1} \cdot \nabla F(\mathbf{x}^i)$$

- But we may not know $\mathbf{H}^{-1}$ or it might be hard to compute.

- Build up estimate for $\mathbf{H}^{-1}$ with successive updates, using information from (many) previous iteration(s) ... various prescriptions, e.g., BFGS (Broyden-Fletcher-Goldfarb-Shanno)

- Related ideas used in density mixing….

Which technique?

Depends on:

- Number of variables (matrix size)

  (determines, e.g., whether one can afford to save information from many iterations)

- Shape of surface $F(\mathbf{x})$
  - how close to quadratic form
  - how isotropic
  - how sparse $\mathbf{H}$ is
  - extent to which $\mathbf{H}$ is diagonally dominant.

Kohn-Sham as minimization problem

- $\mathbf{x} \leftrightarrow \psi$ or $\{c_i\}$ (plane wave coefficients)

- $F \leftrightarrow E$ (Kohn-Sham energy functional) , subject to orthonormality constraints:

$$\langle \psi_b | \psi_b' \rangle = \delta_{b,b'}$$

- $-\nabla F(\mathbf{x}) \leftrightarrow \mathbf{H}\psi +$ *terms from orthonormality constraints*

- Band b + normalization constraint $\rightarrow$ - $\nabla F(\mathbf{x}) \leftrightarrow H\psi_b - \epsilon_b\psi_b$

- Usually, orthogonalize in separate steps (Gram-Schmidt, etc.)

- Note, for self-consistency, $\mathbf{H}$ updated during minimization steps.

Car-Parrinello Method

- Molecular dynamics way of simultaneously doing ionic dynamics and solving electronic problem.

- Regard each c_i as coordinate of a classical particle.

- Fictitious dynamics of c_i's.

- Solve classical equations of motion to update c_i's.

- Solve the electronic problem by quenched molecular dynamics (e.g., simulated annealing) – give some K.E. then cool, should end up at minimum.

- Need very small time steps to solve electronic problem well.

- May converge to linear combination of the Kohn-Sham eigenstates, need to do a "subspace diagonalization" to get the correct ones.

R. Car and M. Parrinello, Phys. Rev. Lett. 55 2471 (1985).

Direct Minimization: Conjugate Gradient

Either do one band at a time or can do
  all bands simultaneously.

- Initial guess $\{c_i\}^1$
- Calculate density
- Calculate potential $V_{eff}(\mathbf{r})$
- Calculate gradient
- If gradient small, stop
- Calculate (preconditioned) conjugate direction
- Calculate new $\{c_i\}$
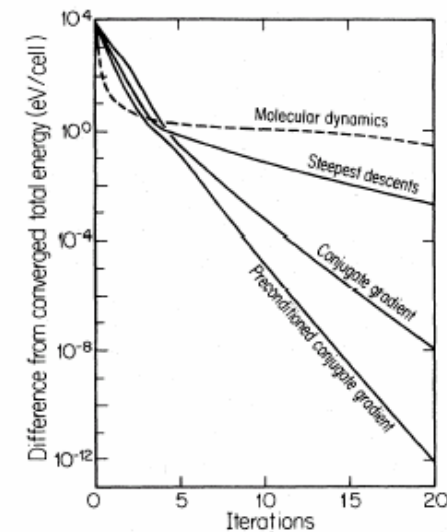- Orthogonalize $\{c_i\}$

FIG. 18. Error in the total energy of an 8-atom silicon supercell with a 32-Ry kinetic-energy cutoff vs iteration number for *indirect-minimization* (dashed line) and *direct-minimization* (solid line) methods. Note that the curve labeled "molecular dynamics" involves a first-order equation of motion, and the number of iterations associated with this curve *has been divided by five* to allow comparison at the same level of computational effort as discussed in the text.

Review: Payne, Teter, Allan, Arias & Joannopoulos, Rev. Mod. Phys. 64 1045 (1992).

Iterative Diagonalizers -1
Krylov Subspaces

- Recall in conjugate gradient, the set of search directions $\{\mathbf{d}^i\}$, $i=1,n$ was increased at each step by the addition of a new direction $\mathbf{d}^{n+1}$ which is obtained from $\mathbf{H} \cdot \mathbf{x}^n$

- Krylov subspace of a matrix $\mathbf{A}$: $(\mathbf{Ax}, \mathbf{A}^2\mathbf{x}, \mathbf{A}^3\mathbf{x}, \dots)$

- Way of increasing the space in which one searches for a solution.

- Note: Unless $\mathbf{x}$ is an eigenvector of $\mathbf{A}$, $\mathbf{Ax}$ will point in a new direction.

Iterative Diagonalizers -2
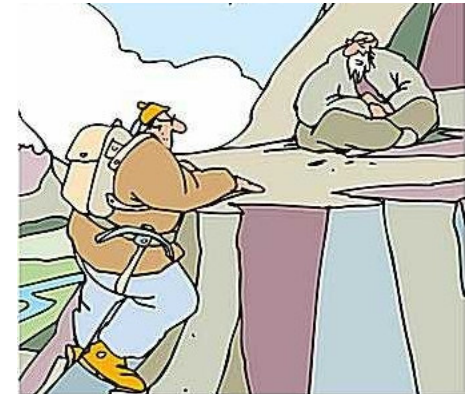# Krylov Subspaces (contd.)

Found on "Google Answers"...

- Question: Why is the Krylov subspace the natural space to look for a solution?

- Answer: Your Question about the relationship between Krylov subspaces and iterative solution of linear systems brings to mind an old story: One dark evening a man sees a friend, busily searching for something under a street lamp. When he asks how he can help, he's told that a watch is missing, and for several minutes both of them diligently look for it. Finally the man asks his friend if he's sure this is where the watch was lost. "Oh no," the friend replies, "I left it in the park, but the light is much better here."

  Certainly it is not obvious that a Krylov subspace would have the solution of a linear system, or even a good estimated solution. However it is easy to search there! And under some conditions the search turns out to be unreasonably successful.

Iterative Diagonalizers -3

- Differences in approaches :
  - How much of the Krylov subspace is used explicitly at each iteration.
  - Way in which previous iterates are combined to give new one.
  - Whether solution is band-by-band or all bands (for Kohn-Sham problem)
- Some popular iterative diagonalizers:
  - Lanczos
  - Conjugate Gradient
  - Davidson

Sparseness & Use of FFT's

- Iterative methods: repeatedly need to calculate $H\psi$.

- If the matrix H is sparse (lots of zero elements), this is fast.

- $H = T + V$.

- Kinetic energy diagonal in $\mathbf{G},\mathbf{G}'$, so $T\psi$ fast in reciprocal space.

- But potential energy is diagonal in $\mathbf{r},\mathbf{r}'$ so $V\psi$ fast in real space.

- Also calculating density $n(\mathbf{r})$ fast in real space.

- Use Fast Fourier Transforms (FFTs) to go back and forth between real space and reciprocal space.

- Cost of FFT $\sim N_{PW}\, ln\, (N_{PW})$

Mixing

- Iterations n of self-consistent cycle:

- Successive approximations to density: $n_{in}(n) \rightarrow n_{out}(n) \rightarrow n_{in}(n+1)$.

- $n_{out}(n)$ fed directly as $n_{in}(n+1)$?? No, usually doesn't converge.

- Need to mix, take some combination of input and output densities (may include information from several previous iterations).

- Goal is to achieve self consistency ($n_{out} = n_{in}$) in as few iterations as possible.
- Want to solve $\mathbf{f}(\mathbf{x})=\mathbf{x} \rightarrow$ Minimize norm of $\mathbf{R}[\mathbf{x}] = \mathbf{f}(\mathbf{x}) - \mathbf{x}$

- Simplest prescription = linear mixing:
$$n_{in}(n+1) = \alpha\, n_{out}(n) + (1-\alpha)\, n_{in}(n).$$

- There exist more sophisticated prescriptions (Broyden mixing, modified Broyden mixing of various kinds...) based on QNR methods.

- Some people mix potentials instead of densities.

Total Energy

- Very useful quantity!

- Can use to get structures, heats of formation, adsorption energies, diffusion barriers, activation energies, elastic moduli, vibrational frequencies,…

- Not just sum of eigenvalues because of double counting of electron-electron terms (Hartree, Exchange, Correlation).

- Also need to add ion-ion interaction (Ewald sums).

- Cancellation of divergences at $\mathbf{G}=0$

Geometry Optimization-1

- Simplest case: only have to vary one degree of freedom
    - e.g., structure of diatomic molecule
    - e.g., lattice constant of a cubic (SC, BCC, FCC) crystal
- Can just look for minimum in binding curve (total energy vs. a)

Ag bulk – E_{tot} vs. a
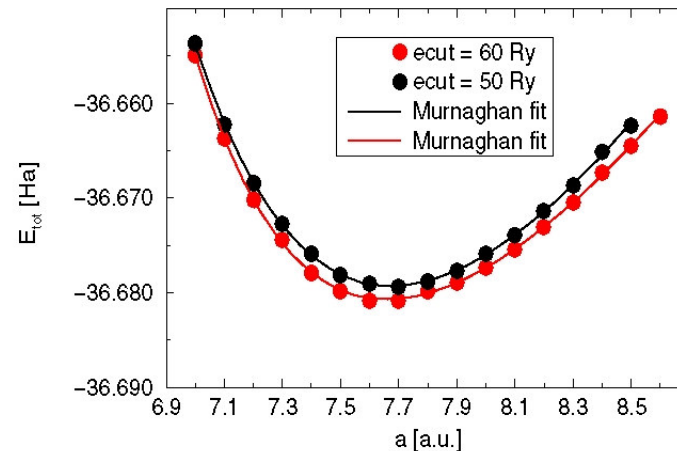
Forces

- Need for geometry optimization and molecular dynamics.
- Could get as finite differences of total energy - too expensive!
- Use force (Hellmann-Feynman) theorem:
  - Want to calculate the force on ion I:

$$\mathbf{F}_I = -\frac{d}{d\mathbf{R}_I}\langle\Psi|H|\Psi\rangle$$

  - Get three terms:

$$\mathbf{F}_I = -\langle\Psi|\frac{\partial H}{\partial\mathbf{R}_I}|\Psi\rangle - \langle\frac{\partial\Psi}{\partial\mathbf{R}_I}|H|\Psi\rangle - \langle\Psi|H|\frac{\partial\Psi}{\partial\mathbf{R}_I}\rangle$$

Wher $|\Psi\rangle$ is an eigenstate, $H|\Psi\rangle = E|\Psi\rangle$

-Substitute this...

Forces (contd.)

- The force is now given by

$$\mathbf{F}_I = -\langle\Psi|\frac{\partial H}{\partial \mathbf{R}_I}|\Psi\rangle - E\langle\frac{\partial \Psi}{\partial \mathbf{R}_I}|\Psi\rangle - E\langle\Psi|\frac{\partial \Psi}{\partial \mathbf{R}_I}\rangle$$

$$= -\langle\Psi|\frac{\partial H}{\partial \mathbf{R}_I}|\Psi\rangle - E\frac{\partial}{\partial \mathbf{R}_I}\langle\Psi|\Psi\rangle$$

- Note that we can now calculate the force from a calculation at ONE configuration alone – huge savings in time.

- If the basis depends upon ionic positions (not true for plane waves), would have extra terms = Pulay forces.

- If $|\Psi\rangle$ is not an exact eigenstate (electronic calculation not well converged) may get big errors in forces calculated using this prescription.

Stress

- Strain: $\mathbf{r}_\alpha \to (\delta_{\alpha\beta} + \epsilon_{\alpha\beta})\mathbf{r}_\beta$

- Stress: $\sigma_{\alpha\beta} = -\frac{1}{\Omega}\frac{\partial E}{\partial \epsilon_{\alpha\beta}}$

- Stress Theorem (Nielsen & Martin, 1985) $\to$
  as for forces, can calculate at a single configuration.

- What if the primitive lattice vectors (specifying unit cell) are not optimal?

  - Forces on atoms may = 0 (e.g., an FCC crystal with wrong lattice constant)

  - Stress will not be zero, however.

    $< 0 \to$ cell would like to expand.

    $> 0 \to$ cell would like to contract.

Geometry Optimization-2

- Conditions for equilibrium:
  - Forces on all atoms = 0.
  - Stress = externally applied stress
- Variables whose values to be optimized:
  - nuclear coordinates $\{\mathbf{R}_I\}$
  - unit cell vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ (6 out of 9 are independent variables)
- Now optimize using any routine of choice.
- Again, various algorithms; one popular one = BFGS.

☠ Symmetries could change during geometry optimization (may or may not be permitted), check k points
Number of plane waves may change discontinuously when cell size changed.

Bibliography

- Martin: "Electronic Structure, Basic Theory & Practical Methods", Cambridge, 2004…and references therein.

- Payne, Teter, Allan, Arias & Joannopoulos, Rev. Mod. Phys. 64 1045, 1992.

- Kresse & Furthmuller, Comput. Mat. Sci. 6, 15, 1996.

- Press, Vetterling, Teukolsky & Flannery, "Numerical Recipes", Cambridge, 1992.

- Shewchuk, "An introduction to the Conjugate Gradient Method Without the Agonizing Pain",

  http://www.cmu.edu/~quake-papers/painless-conjugate-gradient.ps