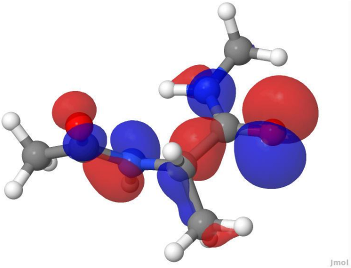
The nuts and bolts of electronic structure theory

Part II: Self-consistency, gradients, relaxation, and vibrations

Jürgen Wieferink

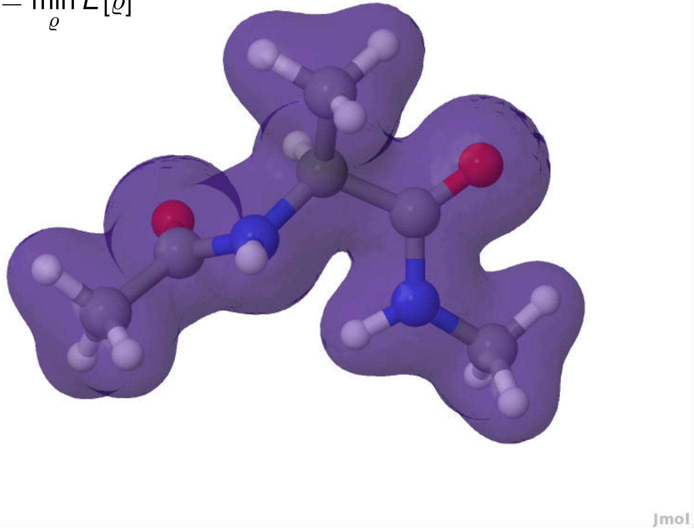
Fritz-Haber-Institut der Max-Planck-Gesellschaft

13 July 2011

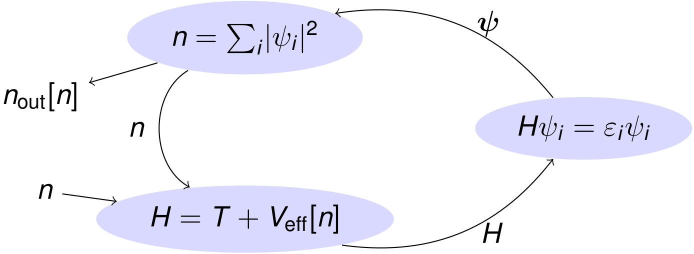


Self-consistency

$$E = \min_{\varrho} E[\varrho]$$



The self-consistency cycle



$n = \sum_i |\psi_i|^2$

$H\psi_i = \varepsilon_i\psi_i$

$H = T + V_{\text{eff}}[n]$

Naive iteration

We want to find the self-consistent solution to

$$n \overset{!}{=} n_{\text{out}}[n],$$

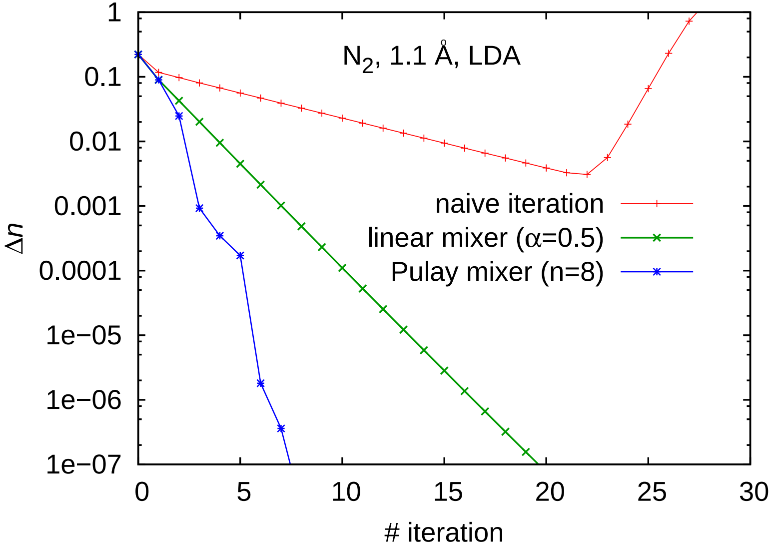
or, equivalently, the zero of the residual

$$0 \overset{!}{=} \Delta n = \Delta n[n] := n_{\text{out}}[n] - n.$$

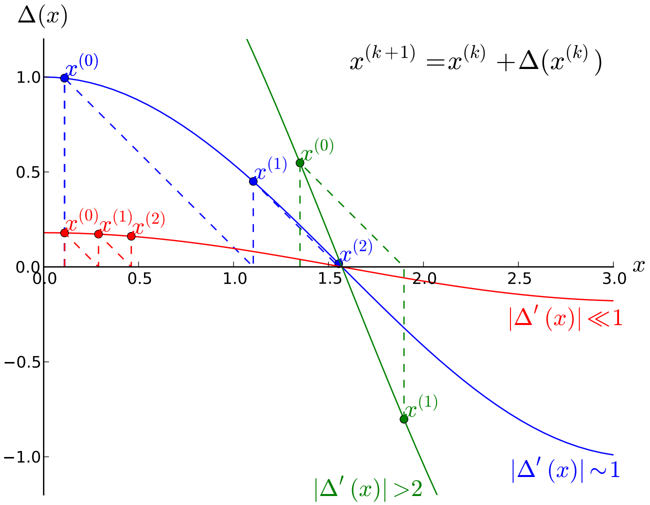
The naive solution is to iterate

$$n^{(k+1)} := n_{\text{out}}^{(k)} = n^{(k)} + \Delta n^{(k)}.$$

Pulay mixer in practice



Naive iteration in one dimension



Linear mixing

- The iterations $x^{(k+1)} = x^{(k)} + \Delta(x^{(k)})$ converge …
  - slow for $|\Delta'| \ll 1$
  - fast for $|\Delta'| \sim 1$
  - not at all for $|\Delta'| > 2$.
- Can always ensure convergence by scaling

$$\Delta(x) \to \alpha\,\Delta(x) \quad \text{with} \quad \alpha < 1$$

  or

$$n^{(k+1)} = n^{(k)} + \alpha\,\Delta n^{(k)} = (1-\alpha)n^{(k)} + \alpha n_{\text{out}}^{(k)}.$$

- But: Problem is multidimensional.
  Different “directions” of Δn would need different α.

Pulay mixer

- After M iterations, we already know about

$$\Delta n^{(i)} = \Delta n[n^{(i)}] \quad \text{for } i = 1, \ldots, M$$

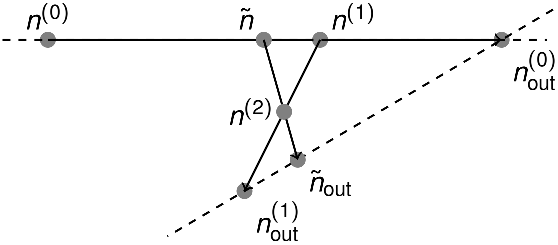
- Linearly approximate residue:

$$\Delta n\Big[\textstyle\sum_i c_i n^{(i)}\Big] \approx \textstyle\sum_i c_i \Delta n^{(i)} =: \Delta n(\mathbf{c})$$

- Minimize $\big\langle \Delta n(\mathbf{c}) \big| \Delta n(\mathbf{c}) \big\rangle$ with respect to $\mathbf{c}$ and $\sum_i c_i = 1$
- “Direct inversion in iterative subspace” (DIIS)
- Other mixers exist and differ in details [2]

[1] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[2] V. Eyert, J. Comp. Phys. **124**, 271 (1996).

Pulay mixer in two dimensions



[1] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).

Idea of electronic smearing

- Calculate charge density from

  $$n(\boldsymbol{r}) = \sum_i \Theta(\mu - \varepsilon_i)|\psi_i(\boldsymbol{r})|^2$$

  with chemical potential μ adjusted to give N_e.
- Tiny changes in $V_\mathrm{eff}[n]$
  → tiny changes in ε_i
  → possibly huge changes in $n(\boldsymbol{r})$.
- Get smooth dependence by smearing $\Theta(\mu - \varepsilon_i)$.

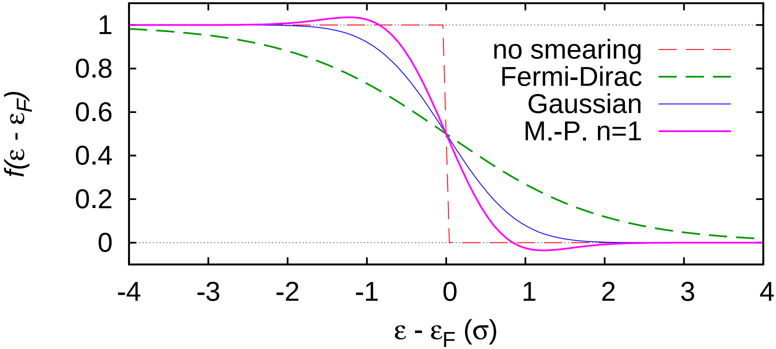
Approaches to electronic smearing

Replace $\Theta(\varepsilon_\mathrm{F} - \varepsilon_i)$ by $f(x)$ with $x = (\varepsilon_i - \mu)$ in:

$$f(x) = \begin{cases} \Theta(-x) & \text{no smearing} \\ \dfrac{1}{e^{\frac{x}{\sigma}} + 1} & \text{Fermi-Dirac [1]} \\ \frac{1}{2}\,\mathrm{erfc}\,\frac{x}{\sigma} & \text{Gaussian [2]} \\ \frac{1}{2}\,\mathrm{erfc}\,\frac{x}{\sigma} + \displaystyle\sum_{m=1}^{N} A_m H_{2m-1}(\tfrac{x}{\sigma}) e^{-\frac{x^2}{\sigma^2}} & \text{Methfessel-Paxton [3]} \end{cases}$$

[1] N. Mermin, Phys. Rev. **137**, A1441 (1965).
[2] C.-L. Fu, K.-H. Ho, Phys. Rev. B **28**, 5480 (1983).
[3] M. Methfessel, A. Paxton, Phys. Rev. B **40**, 3616 (1989).

Comparison of electronic smearing approaches



[1] N. Mermin, Phys. Rev. **137**, A1441 (1965).
[2] C.-L. Fu, K.-H. Ho, Phys. Rev. B **28**, 5480 (1983).
[3] M. Methfessel, A. Paxton, Phys. Rev. B **40**, 3616 (1989).

Electronic “free energy”

- Kohn-Sham equations now from minimizing a “free energy”:

$$E_{\text{tot}} \rightarrow \Omega_{\text{tot}} := E_{\text{tot}} - \sigma S(\sigma, \{f_i\}).$$

- The shape of the entropy term depends on the smearing type.
- The true ($T \rightarrow 0$) total energy is

$$\Omega_{\text{tot}}(\sigma) \leq E_{\text{tot}}(T \rightarrow 0) \leq E_{\text{tot}}(\sigma).$$

  as can be shown by applying variational principle for $\sigma \neq 0$ and $\sigma = T = 0$.

- For metallic systems, $E_{\text{tot}}(T \rightarrow 0)$ can be backextrapolated from $\Omega_{\text{tot}}(\sigma)$ and $E_{\text{tot}}(\sigma)$.

Preconditioner

- Linear admixture of residual:

$$n \leftarrow n + \alpha \, \Delta n$$

often overemphasizes large rearrangements of charge.

- Avoid charge sloshing with preconditioner

$$n \leftarrow n + \hat{G} \, \Delta n \quad \text{with} \quad \hat{G} = \alpha \frac{k^2}{k^2 + k_0^2}$$

which damps long-range ($k \rightarrow 0$) components.

[1] G. Kerker, Phys. Rev. B **23**, 3082 (1981).

Summary self-consistency

- Cast Kohn-Sham minimization to self-consistency problem
- Linear mixing guarantees convergence (but slow)
- Pulay mixer linearly approximates residues
- Electronic smearing to soften discontinuities
- Preconditioner dampens long-range part of residue

Gradients

$$\boldsymbol{F} = -\boldsymbol{\nabla} E(\boldsymbol{R})$$

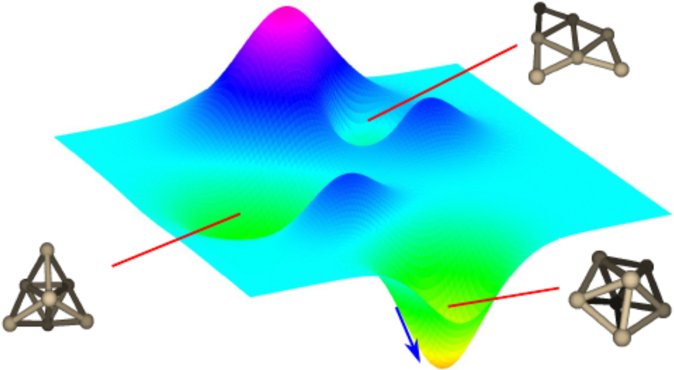


figure adapted from R. Gehrke

Separation of dependencies

Atomic motion means concomitant motion of several quantities:

- the centers $\boldsymbol{R}_{\text{pot}}$ of the external potential $^{Z_\mu}/_{|\boldsymbol{r}-\boldsymbol{R}_{\text{pot},\mu}|}$,
- the centers $\boldsymbol{R}_{\text{basis}}$ of the basis functions $\phi_{nlm}(\boldsymbol{r} - \boldsymbol{R}_{\text{basis},\mu})$,
- a few other centers for particular approximations

Of course, they are *always* the same

$$\boldsymbol{R}_{\text{pot}}(\boldsymbol{R}) = \boldsymbol{R}_{\text{basis}}(\boldsymbol{R}) = \cdots = \boldsymbol{R}.$$

Chain rule

Although all centers move concomitantly

$$\boldsymbol{R}_{\text{pot}}(\boldsymbol{R}) = \boldsymbol{R}_{\text{basis}}(\boldsymbol{R}) = \cdots = \boldsymbol{R}$$

we can use them to separate the gradients into independent terms

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}\boldsymbol{R}} &= \frac{\partial E}{\partial \boldsymbol{R}_{\text{pot}}} \overbrace{\frac{\partial \boldsymbol{R}_{\text{pot}}}{\partial \boldsymbol{R}}}^{=1} + \frac{\partial E}{\partial \boldsymbol{R}_{\text{basis}}} \overbrace{\frac{\partial \boldsymbol{R}_{\text{basis}}}{\partial \boldsymbol{R}}}^{=1} + \cdots \\ &= \frac{\partial E}{\partial \boldsymbol{R}_{\text{pot}}} + \frac{\partial E}{\partial \boldsymbol{R}_{\text{basis}}} + \cdots\end{aligned}$$

Hellmann-Feynman forces

$$\frac{\mathrm{d}E}{\mathrm{d}\boldsymbol{R}_{\text{pot}}} = \frac{\partial E}{\partial \boldsymbol{R}_{\text{pot}}} + \overbrace{\sum_{\alpha i} \frac{\partial E}{\partial c_{\alpha i}} \frac{\partial c_{\alpha i}}{\partial \boldsymbol{R}_{\text{pot}}}}^{=:0}$$

- E is minimal w. r. t. changes in one-particle coefficients $c_{\alpha i}$ consistent with orthonormality.
- Orthonormality defined by $\boldsymbol{R}_{\text{basis}}$, not $\boldsymbol{R}_{\text{pot}}$.

$$\begin{aligned} \boldsymbol{F}_{\text{HF},\mu} &= -\frac{\partial E_{\text{nuc-nuc}}}{\partial \boldsymbol{R}_{\text{pot},\mu}} - \frac{\partial E_{\text{ext}}}{\partial \boldsymbol{R}_{\text{pot},\mu}} \\ &= -\sum_{\nu} Z_\nu \boldsymbol{\nabla}_{\boldsymbol{R}_\mu} \frac{Z_\mu}{|\boldsymbol{R}_\nu - \boldsymbol{R}_\mu|} + \int \mathrm{d}^3 r \; n(\boldsymbol{r}) \boldsymbol{\nabla}_{\boldsymbol{R}_\mu} \frac{Z_\mu}{|\boldsymbol{r} - \boldsymbol{R}_\mu|} \end{aligned}$$

Pulay forces

$$\frac{\mathrm{d}E}{\mathrm{d}\boldsymbol{R}_{\text{basis}}} = \sum_{\alpha} \frac{\partial E}{\partial \phi_\alpha} \frac{\partial \phi_\alpha}{\partial \boldsymbol{R}_{\text{basis}}} + \sum_{\alpha i} \frac{\partial E}{\partial c_{\alpha i}} \frac{\partial c_{\alpha i}}{\partial \boldsymbol{R}_{\text{basis}}}$$

- None of these terms vanish.
- Make use of
  - definition of H,
  - Kohn-Sham equations,
  - orthonormality.

$$\boldsymbol{F}_{\text{Pulay},\mu} = -2 \sum_{\alpha,\alpha',i} c^*_{\alpha i} \left\langle \frac{\partial \phi_\alpha}{\partial \boldsymbol{R}_{\text{basis},\mu}} \middle| H - \varepsilon_i \middle| \phi_{\alpha'} \right\rangle c_{\alpha' i}$$

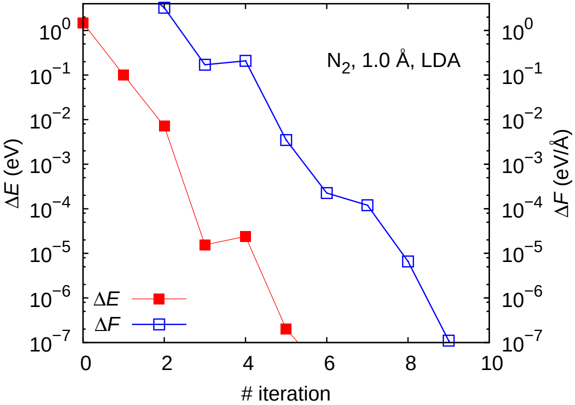
- Particularly involved for GGAs

Other force terms

- Moving integration grid $\boldsymbol{R}_{\text{grid}}$:
  - Force term vanishes for converged grid settings.
  - Can be done, but is cumbersome.
  - If really needed, the grid is too coarse anyway [1].
- Electrostatic multipole derivatives (Hartree potential):
  - Important but straight-forward.
- Scalar relativistic corrections:
  - Important but straight-forward.

[1] Baker *et al.*, J. Chem. Phys. **101**, 8894 (1994).

Practitioner's guide



Forces converge slower (first order) within the SCF procedure than energies (second order).

Summary gradients

- Forces are invaluable probe for potential energy surface
- Forces separate in distinct terms:
    - Hellmann-Feynman from moving potential centers
    - Pulay from moving basis functions
    - additional terms from other moving approximations.
- Forces need better SCF convergence than energies.

Relaxation

$$E_{\text{min}} = \min_{\boldsymbol{R}} E(\boldsymbol{R})$$

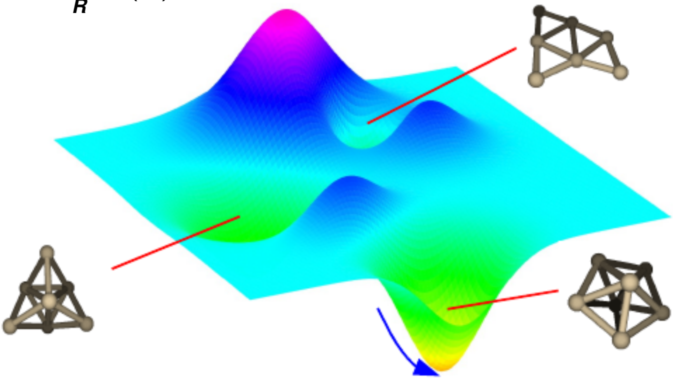


figure adapted from R. Gehrke

Steepest descent

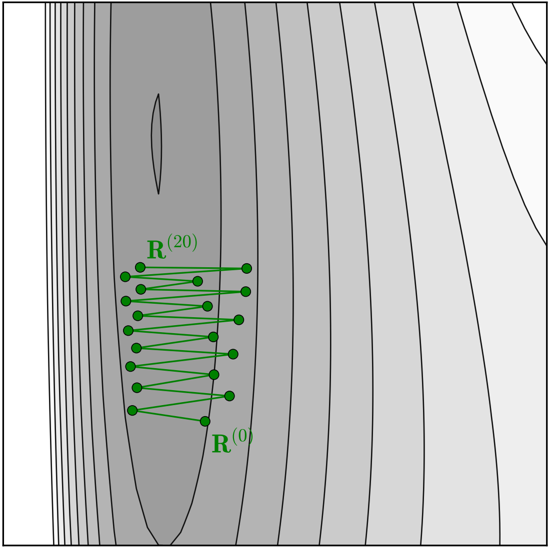
- Minimum of energy is zero of force:

$$\min_{\boldsymbol{R}} E(\boldsymbol{R}) \quad \Rightarrow \quad \boldsymbol{F}(\boldsymbol{R}) \overset{!}{=} 0.$$

- Similar to $\Delta n \overset{!}{=} 0$ in electronic self-consistency.
- E. g. linear mixing corresponds to “steepest descent”

$$\boldsymbol{R}^{(k+1)} := \boldsymbol{R}^{(k)} + \alpha \boldsymbol{F}\big(\boldsymbol{R}^{(k)}\big).$$

Steepest descent in practice

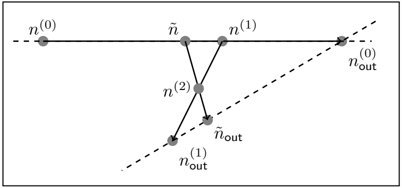


Step in direction of steepest descent:

$$\Delta \boldsymbol{R} = \alpha \boldsymbol{F}$$

Geometrical DIIS

- The Pulay mixer corresponds to “Geometrical direct inversion in iterative subspace” (GDIIS) [1].

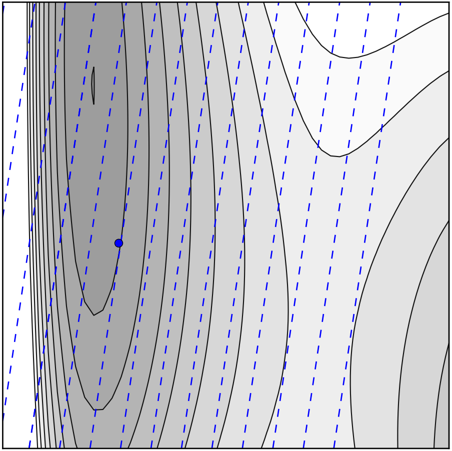


- Special precautions to avoid problems with saddle points [2].
- Very low administrative footprint.

[1] P. Császár and P. Pulay, J. Mol. Struct. **114**, 31 (1984).
[2] Ö. Farkas and H. B. Schlegel, Phys. Chem. Chem. Phys. **4**, 11 (2002).

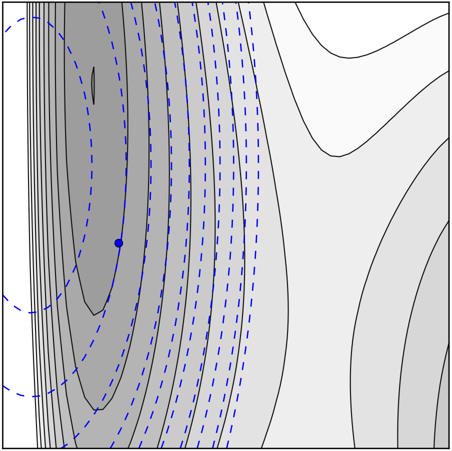
Quadratic model

$$\tilde{E}(\boldsymbol{R} + \Delta\boldsymbol{R}) = E(\boldsymbol{R}) - \boldsymbol{F}(\boldsymbol{R}) \cdot \Delta\boldsymbol{R}$$



Quadratic model

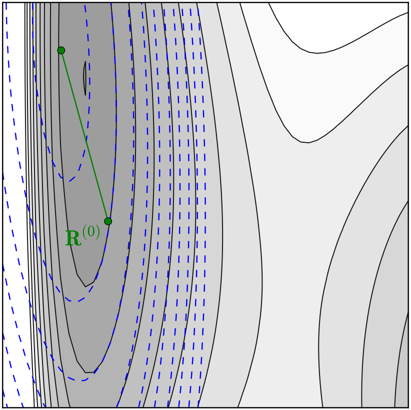
$$\tilde{E}(\boldsymbol{R} + \Delta\boldsymbol{R}) = E(\boldsymbol{R}) - \boldsymbol{F}(\boldsymbol{R}) \cdot \Delta\boldsymbol{R} + \frac{1}{2}\Delta\boldsymbol{R}^T \, {\color{red}\tilde{\mathbf{H}}} \, \Delta\boldsymbol{R}$$



(Quasi-)Newton step

Step to model minimum $\boldsymbol{F}(\boldsymbol{R}) - \tilde{\mathbf{H}}\,\Delta\boldsymbol{R} \overset{!}{=} 0 \quad \Rightarrow \quad \Delta\boldsymbol{R} = \tilde{\mathbf{H}}^{-1}\boldsymbol{F}$

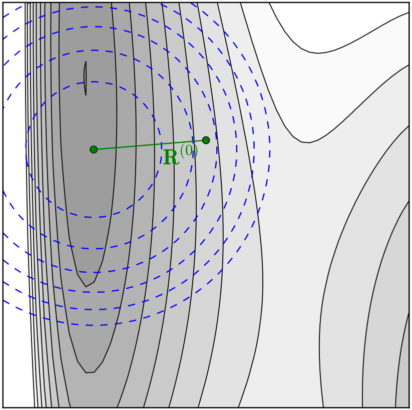
Newton step



Exact Hessian

$\tilde{\mathbf{H}} = \mathbf{H}$

Quasi-Newton step



Model Hessian

$\tilde{\mathbf{H}} = \beta\mathbf{1}$

Hessian approximation

Twofold strategy for Hessian approximation

1. Need initial approximation. However, even $\tilde{\mathbf{H}}_0 = \beta \mathbf{1}$ is not so bad.
2. After each step, update Hessian to fulfill

$$\tilde{\mathbf{H}}\,\Delta \boldsymbol{R} = -\Delta \boldsymbol{F},$$

e. g. by Broyden-Fletcher-Goldfarb-Shanno (BFGS) update [1]

$$\tilde{\mathbf{H}} \quad \leftarrow \quad \tilde{\mathbf{H}} - \frac{\tilde{\mathbf{H}}\,\Delta \boldsymbol{R}\,(\tilde{\mathbf{H}}\,\Delta \boldsymbol{R})^T}{\Delta \boldsymbol{R}^T\,\tilde{\mathbf{H}}\,\Delta \boldsymbol{R}} - \frac{\Delta \boldsymbol{F}\,\Delta \boldsymbol{F}^T}{\Delta \boldsymbol{F}^T\,\Delta \boldsymbol{R}}.$$

[1] J. Nocedal and S. J. Wright, “Numerical optimization” (Springer, 2006).

Initial Hessian approximation

Diagonal matrix ($\tilde{\mathbf{H}} = \beta\mathbf{1}$): penalty on Cartesian shifts

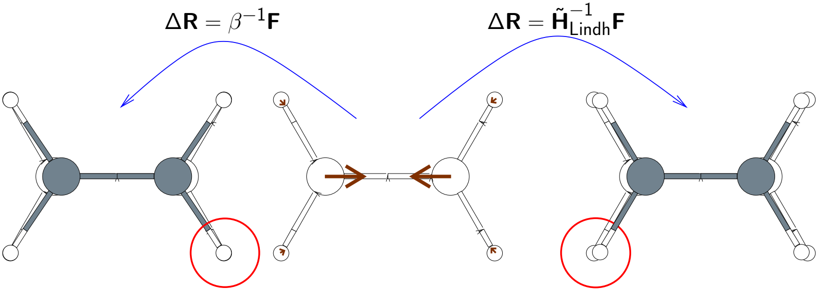
$$\tilde{E}(\boldsymbol{R} + \Delta\boldsymbol{R}) = E - \boldsymbol{F} \cdot \Delta\boldsymbol{R} + \underbrace{\sum_{\nu} \frac{\beta}{2} \Delta R_{\nu}^{2}}_{\substack{\text{Cartesian}\\\text{shifts}}}$$

Lindh [1]: penalty on changes in internal coordinates

$$\tilde{E} = E - \boldsymbol{F} \cdot \Delta\boldsymbol{R} + \underbrace{\sum{}' \frac{k_{ij}}{2} \Delta r_{ij}^{2}}_{\text{stretches}} + \underbrace{\sum{}' \frac{k_{ijk}}{2} \Delta\theta_{ijk}^{2}}_{\text{bendings}} + \underbrace{\sum{}' \frac{k_{ijkl}}{2} \Delta\tau_{ijkl}^{2}}_{\text{torsions}}$$

[1] R. Lindh *et al.*, Chem. Phys. Lett. **241**, 423 (1995).

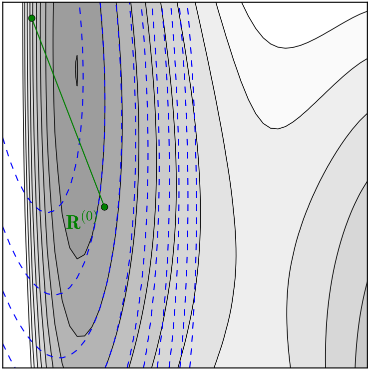
Lindh model in practice



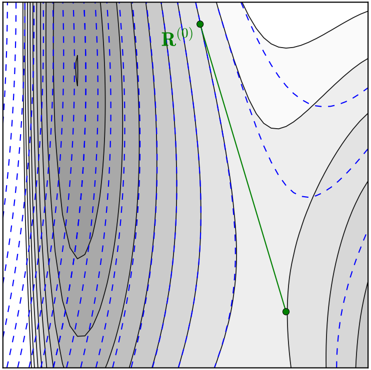
No forces on hydrogen atoms:

- Diagonal matrix preserves Cartesian coordinates of H.
- Lindh matrix preserves bond lengths/angles of H

Failure of pure quasi-Newton



Long steps
$\Delta E > 0$



Negative Hessian mode
Aims at saddle

Step control: line search

$$\Delta \boldsymbol{R} := \alpha \tilde{\mathbf{H}}^{-1} \mathbf{F}$$

- Try Quasi-Newton ($\alpha = 1$) first
- For bad step (e. g. $\Delta E > 0$), interpolate

  $$E(\alpha) = E(\boldsymbol{R} + \alpha\, \Delta \boldsymbol{R})$$

  and find optimal α in 1D
- After each line search, update Hessian
- Drawbacks:
  - Need positive definite Hessian ($\boldsymbol{u}^T \tilde{\mathbf{H}} \boldsymbol{u} > 0\ \forall \boldsymbol{u}$)
  - Might need multiple steps in same direction
- Can avoid explicit $\tilde{\mathbf{H}}$ with L-BFGS.

[1] J. Nocedal and S. J. Wright, “Numerical optimization” (Springer, 2006).

Step control: trust radius

$$\Delta \boldsymbol{R} := \operatorname*{arg\,min}_{|\Delta \boldsymbol{R}| \leq \Delta R_{\mathrm{max}}} \tilde{E}(\boldsymbol{R} + \Delta \boldsymbol{R})$$

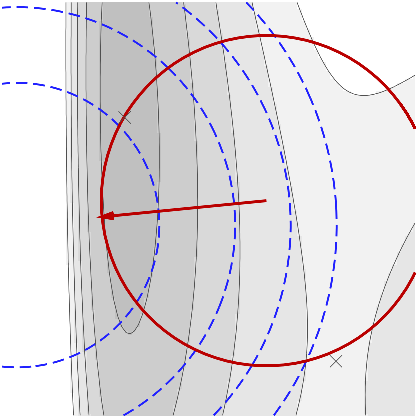
- Only trust quadratic model within finite region $|\Delta \boldsymbol{R}| \leq \Delta R_{\mathrm{max}}$.
- Model minimum can be calculated exactly.
- Initial trust radius can be guessed (e. g. $\Delta R_{\mathrm{max}} = 0.2\,\text{Å}$).
- Update trust radius according to

$$\rho = \frac{\Delta E}{\Delta \tilde{E}} \begin{cases} > 0.75 & \Delta R_{\mathrm{max}} \leftarrow 2 \Delta R_{\mathrm{max}} \quad \text{(good model)} \\ < 0.25 & \Delta R_{\mathrm{max}} \leftarrow \frac{1}{2} \Delta R_{\mathrm{max}} \quad \text{(bad model)} \end{cases}$$

- Stick to old $\boldsymbol{R}$ if $\rho < 0$.

[1] J. Nocedal and S. J. Wright, “Numerical optimization” (Springer, 2006).

Trusted step



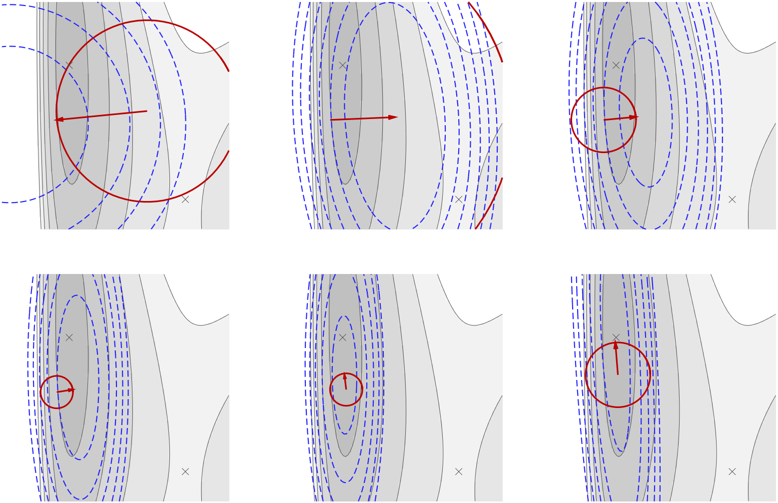
Step to minimum
within trust region:

$$\Delta \boldsymbol{R} := \underset{|\Delta \boldsymbol{R}| \leq \Delta R_{\text{max}}}{\arg\min} \tilde{E}(\boldsymbol{R} + \Delta \boldsymbol{R})$$

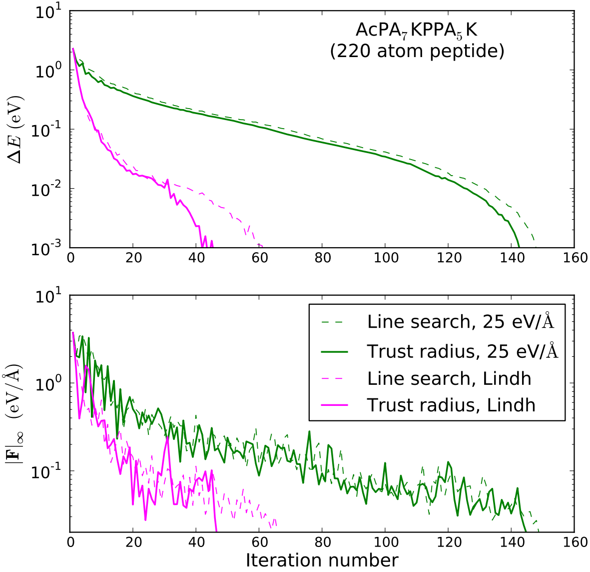
Approximate Hessian

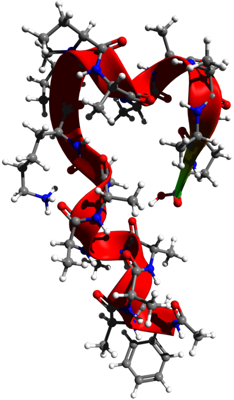
$$\tilde{\mathbf{H}} = \beta \mathbf{1}$$

Trust region method



Example performance





Summary relaxation

- Quasi-Newton methods are “industry standard”
- If optimization runs smooth, use quasi-Newton step
- Methods of step size control:
  - line search
  - trust region
- Hessian approximation:
  - initial model
  - updates to enfoce $\tilde{\mathbf{H}}\,\Delta\boldsymbol{R} = -\Delta\boldsymbol{F}$
- Initial Hessian is decisive for performance

Vibrations

$$\mathbf{M}\,\Delta\ddot{\boldsymbol{R}} = \boldsymbol{F}(\Delta\boldsymbol{R}) \approx -\mathbf{H}\,\Delta\boldsymbol{R}$$

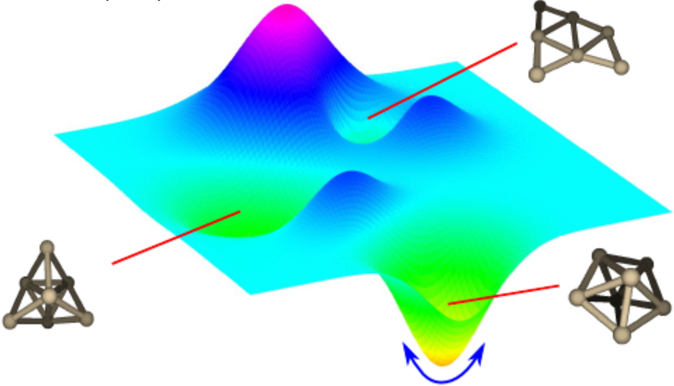


figure adapted from R. Gehrke

Beyond frozen nuclei

Minimum of potential energy surface is ground state given

- classical nuclei and
- zero temperature (no kinetic energy).

Go beyond these approximations to leading order by

$$E(\boldsymbol{R}) \approx \tilde{E}(\boldsymbol{R}) = E_{\text{min}} + \frac{1}{2}\Delta\boldsymbol{R}^T\,\mathbf{H}\,\Delta\boldsymbol{R}$$

with $\Delta\boldsymbol{R} = \boldsymbol{R} - \boldsymbol{R}_{\text{min}}$.

Calculating second derivatives

- Effort for first derivatives (forces $\boldsymbol{F}$) is moderate because energy is stationary w. r. t. electronic structure.
- Analytical second derivatives (Hessian **H**) need second derivative w. r. t. electronic structure.
- Can be done by perturbation theory (see, e. g. [1]), but is cumbersome.
- Alternative: Numerical differentiation:

$$H_{\alpha\beta} = \frac{\mathrm{d}^2 E}{\mathrm{d}R_\alpha \mathrm{d}R_\beta} \approx -\frac{\boldsymbol{F}_\alpha(\boldsymbol{R} + \Delta \boldsymbol{e}_\beta) - \boldsymbol{F}_\alpha(\boldsymbol{R} - \Delta \boldsymbol{e}_\beta)}{2\Delta}$$

[1] S. Baroni *et al.*, Rev. Mod. Phys. **73**, 515 (2001).

Harmonic vibrations

- Newton's equation

$$\mathbf{M}\,\Delta\ddot{\boldsymbol{R}} \overset{!}{=} \tilde{\boldsymbol{F}} = -\mathbf{H}\,\Delta\boldsymbol{R}$$

- Exponential ansatz

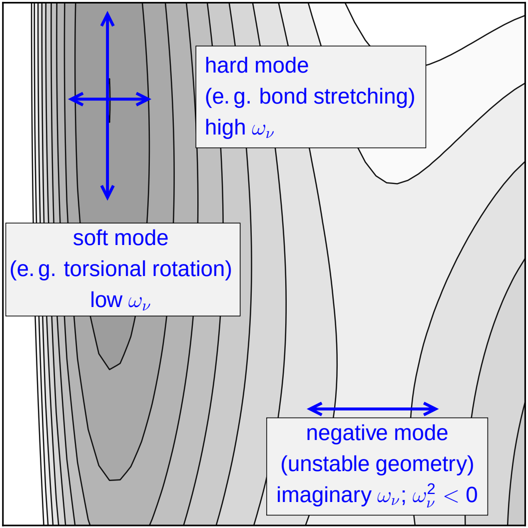
$$\Delta\boldsymbol{R} = \boldsymbol{u}\mathrm{e}^{-\mathrm{i}\omega t} \quad \Rightarrow \quad -\omega^2\mathbf{M}\boldsymbol{u} = -\mathbf{H}\boldsymbol{u}$$

- Generalized eigenproblem $\mathbf{H}\boldsymbol{u}^{(\nu)} = \omega_\nu^2\mathbf{M}\boldsymbol{u}^{(\nu)}$

$$\boldsymbol{R}(t) = \boldsymbol{R}_{\mathrm{min}} + \sum_\nu c_\nu \boldsymbol{u}^{(\nu)} \cos(\omega_\nu t + \Phi_\nu).$$

- Vibrational modes $\boldsymbol{u}^{(\nu)}$ with frequencies ω_ν
- Amplitudes c_ν and phases Φ_ν from initial conditions

Interpretation of modes



hard mode
(e. g. bond stretching)
high ω_ν

soft mode
(e. g. torsional rotation)
low ω_ν

negative mode
(unstable geometry)
imaginary ω_ν; $\omega_\nu^2 < 0$

Vibrational free energy

- Classical solution

$$\boldsymbol{R}(t) = \boldsymbol{R}_{\text{min}} + \sum_{\nu} c_{\nu} \boldsymbol{u}^{(\nu)} \cos(\omega_{\nu} t + \Phi_{\nu}).$$

  is set of independent harmonic oscillators.

- Free energy at T with quantum vibrational effects

$$F(T) = E_{\text{min}} + \sum_{\nu} \left\{ \frac{\hbar\omega_{\nu}}{2} + k_{\text{B}} T \ln\left[1 - e^{-\frac{\hbar\omega_{\nu}}{k_{\text{B}} T}}\right] \right\}$$

- Zero-point vibrational energy

$$E(T=0) = F(T=0) = E_{\text{min}} + \sum_{\nu} \frac{\hbar\omega_{\nu}}{2}$$

Outlook on Molecular Dynamics

- For anharmonic effects we need to go beyond the harmonic approximation.
- Solve Newton's equations

$$\mathbf{M}\ddot{\boldsymbol{R}} = \boldsymbol{F}(\boldsymbol{R})$$

for actual DFT forces $\boldsymbol{F}(\boldsymbol{R})$ as initial value problem

$$\boldsymbol{R}(t=0) = \boldsymbol{R}_0, \qquad \dot{\boldsymbol{R}}(t=0) = \boldsymbol{V}_0$$

- Can also be solved iteratively (e. g. Verlet algorithm)

$$\boldsymbol{R}(t+\Delta t) = 2\boldsymbol{R}(t) - \boldsymbol{R}(t-\Delta t) + \ddot{\boldsymbol{R}}(t)\cdot\Delta t^2 + \mathcal{O}(\Delta t^4)$$

Summary vibrations

- First order correction to classical geometric ground state
- Hessian harder than forces; often resort to finite differences
- Soft, hard, and negative modes
- Free energy and zero-point corrections

Final summary

- Self-consistency
  - → Finding the electronic ground state for a given geometry
- Gradients
  - → Change of energy for infinitesimal changes of geometry
- Relaxation
  - → Finding minimal energy geometries
- Vibrations
  - → Characterizing minimal energy geometries

Thank you for your attention!