Making Electronic Structure Theory work

Ralf Gehrke

Fritz-Haber-Institut der Max-Planck-Gesellschaft
Berlin, 23th June 2009

Hands-on Tutorial on *Ab Initio* Molecular Simulations:
Toward a First-Principles Understanding of Materials Properties and Functions

MAX-PLANCK-GESELLSCHAFT

Outline

- **Achieving Self-Consistency**
  - Density Mixing
    - Linear Mixing, Pulay Mixer, Broyden Mixer
  - Electronic Smearing
    - Fermi-Dirac, Gaussian, Methfessel-Paxton
  - Preconditioning
  - Direct Minimization
- Spin Polarization
- Energy Derivatives
- Local Atomic Structure Optimization
- Outlook on Global Structure Optimization

Outline

- Achieving Self-Consistency
- **Spin Polarization**
  - ➜ Fixed Spin vs. Unconstrained Spin Calculation
- Energy Derivatives
- Local Atomic Structure Optimization
- Outlook on Global Structure Optimization

Outline

- Achieving Self-Consistency
- Spin Polarization
- **Energy Derivatives**
  - ➜ Atomic Forces
  - ➜ Numeric vs. Analytic Second Derivative
  - ➜ Vibrational Frequencies
- Local Atomic Structure Optimization
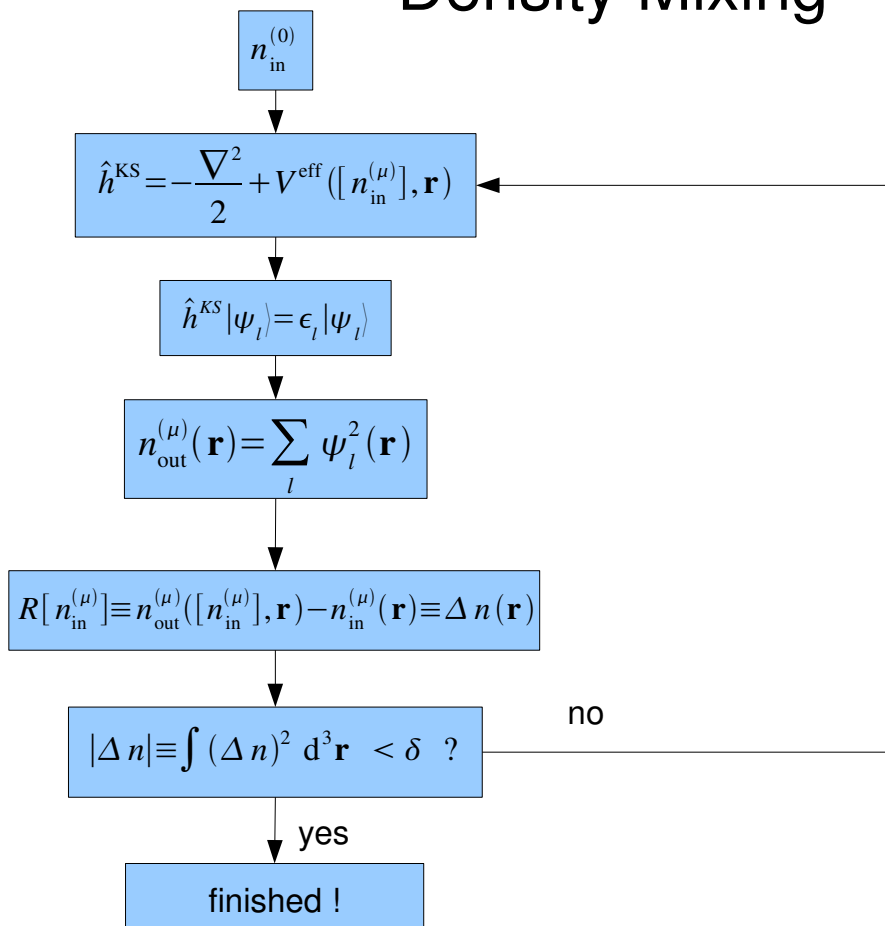- Outlook on Global Structure Optimization

Outline

- Achieving Self-Consistency
- Spin Polarization
- Energy Derivatives
- **Local Atomic Structure Optimization**
  - Steepest Descent
  - Conjugate Gradient
  - Broyden-Fletcher-Goldfarb-Shannon (BFGS)
  - Cartesian vs. Internal Coordinates
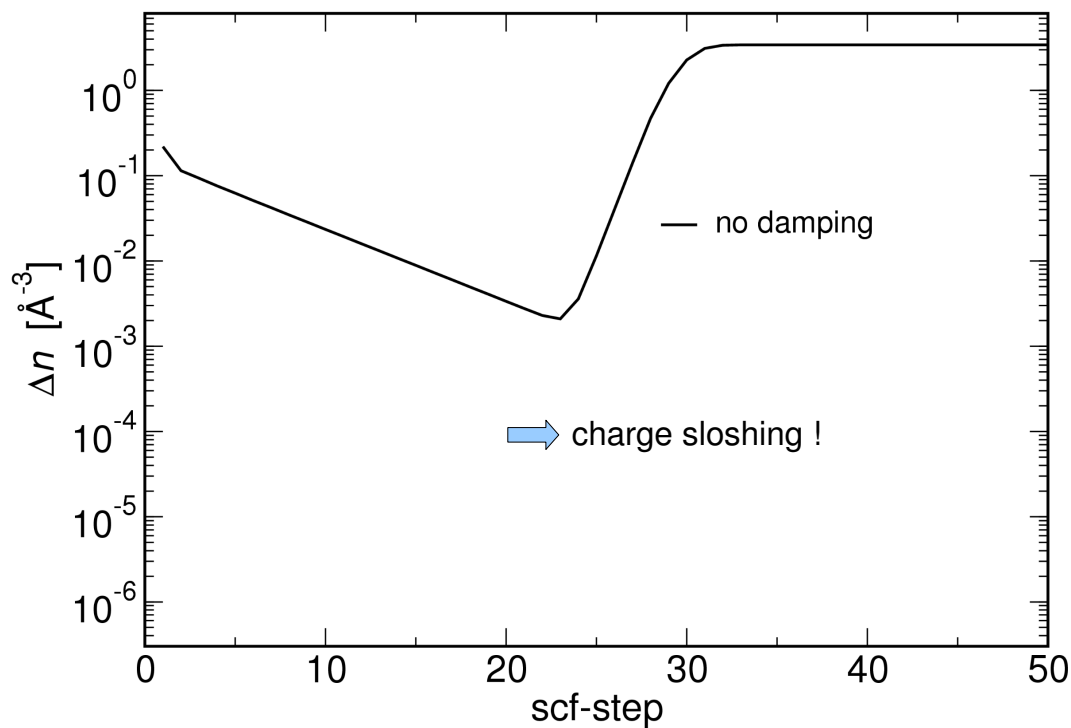- Outlook on Global Structure Optimization

Outline

- Achieving Self-Consistency
- Spin Polarization
- Energy Derivatives
- Local Atomic Structure Optimization
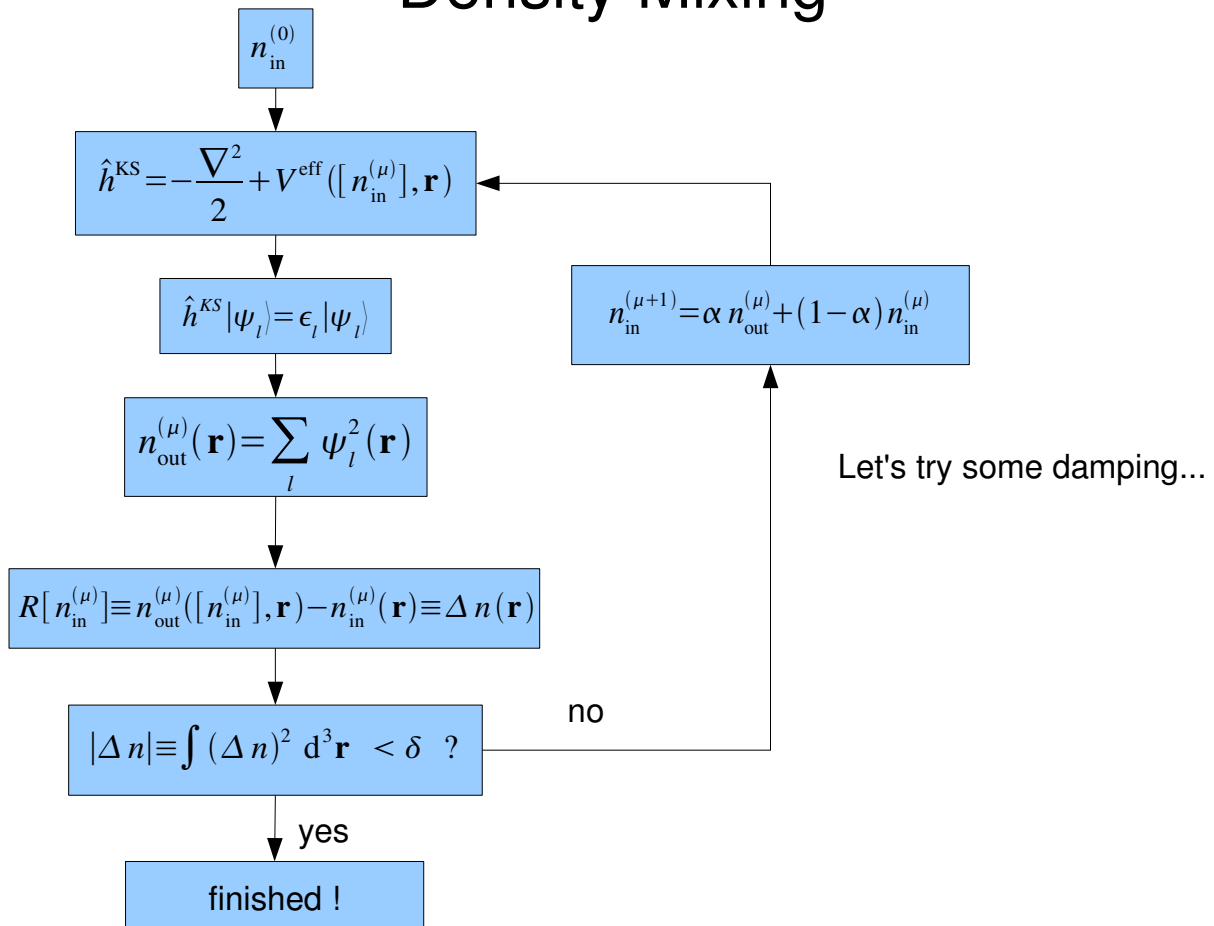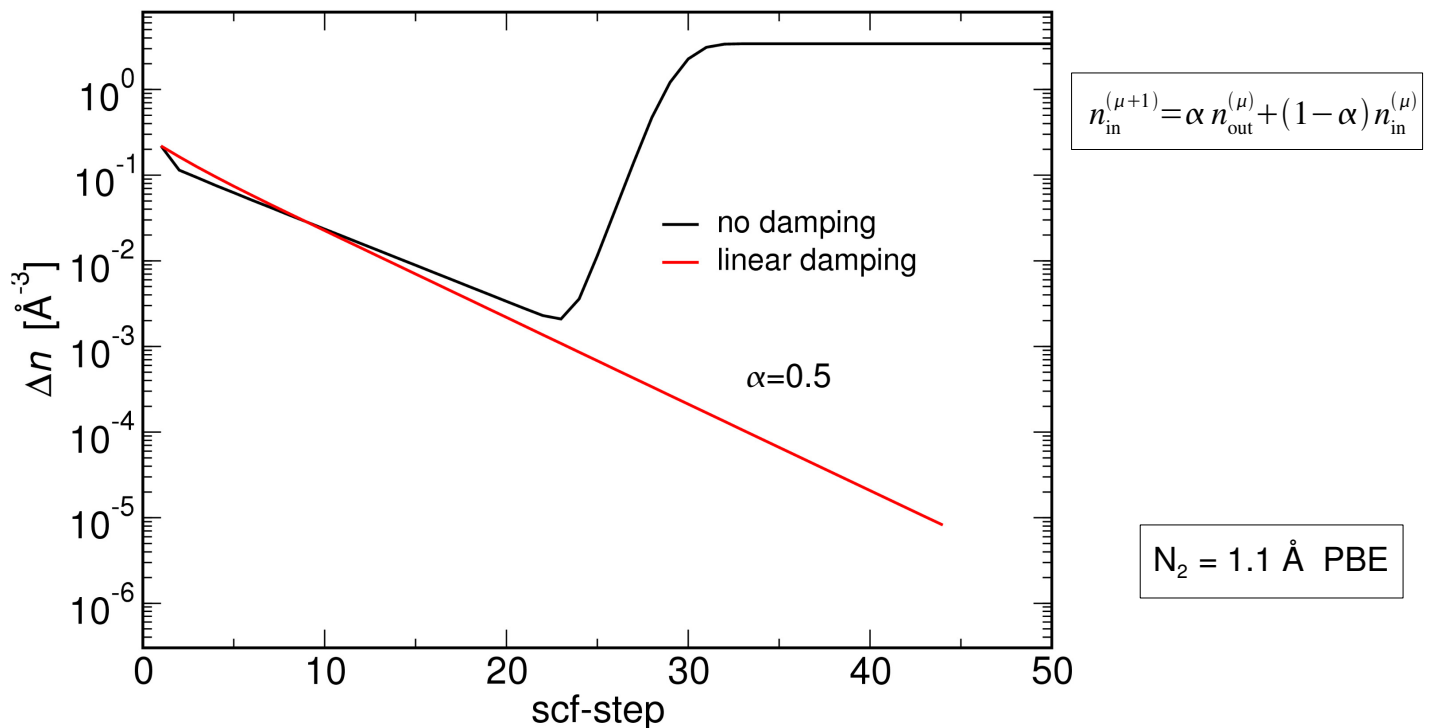- **Outlook on Global Structure Optimization**

Density Mixing

$$n_{\text{in}}^{(0)}$$

$$\hat{h}^{\text{KS}} = -\frac{\nabla^2}{2} + V^{\text{eff}}([n_{\text{in}}^{(\mu)}], \mathbf{r})$$

$$\hat{h}^{\text{KS}} |\psi_l\rangle = \epsilon_l |\psi_l\rangle$$

$$n_{\text{out}}^{(\mu)}(\mathbf{r}) = \sum_l \psi_l^2(\mathbf{r})$$

$$R[n_{\text{in}}^{(\mu)}] \equiv n_{\text{out}}^{(\mu)}([n_{\text{in}}^{(\mu)}], \mathbf{r}) - n_{\text{in}}^{(\mu)}(\mathbf{r}) \equiv \Delta n(\mathbf{r})$$

$$|\Delta n| \equiv \int (\Delta n)^2 \, \mathrm{d}^3\mathbf{r} \ < \delta \ ?$$

no

yes

finished !

Density Mixing

— no damping

⇨ charge sloshing !

$$n_{\text{in}}^{(\mu+1)} = n_{\text{out}}^{(\mu)}$$

$N_2 = 1.1$ Å PBE

Density Mixing

$$n_{in}^{(0)}$$

$$\hat{h}^{KS} = -\frac{\nabla^2}{2} + V^{eff}([n_{in}^{(\mu)}], \mathbf{r})$$

$$\hat{h}^{KS} |\psi_l\rangle = \epsilon_l |\psi_l\rangle$$

$$n_{out}^{(\mu)}(\mathbf{r}) = \sum_l \psi_l^2(\mathbf{r})$$

$$R[n_{in}^{(\mu)}] \equiv n_{out}^{(\mu)}([n_{in}^{(\mu)}], \mathbf{r}) - n_{in}^{(\mu)}(\mathbf{r}) \equiv \Delta n(\mathbf{r})$$

$$|\Delta n| \equiv \int (\Delta n)^2 \, \mathrm{d}^3\mathbf{r} \; < \delta \; ?$$

no

yes

finished !

$$n_{in}^{(\mu+1)} = \alpha \, n_{out}^{(\mu)} + (1-\alpha) \, n_{in}^{(\mu)}$$

Let's try some damping...

Density Mixing

$$n_{in}^{(\mu+1)} = \alpha \, n_{out}^{(\mu)} + (1-\alpha) \, n_{in}^{(\mu)}$$

— no damping
— linear damping

$\alpha = 0.5$

$N_2 = 1.1$ Å PBE

⟹ seems to work !

Density Mixing
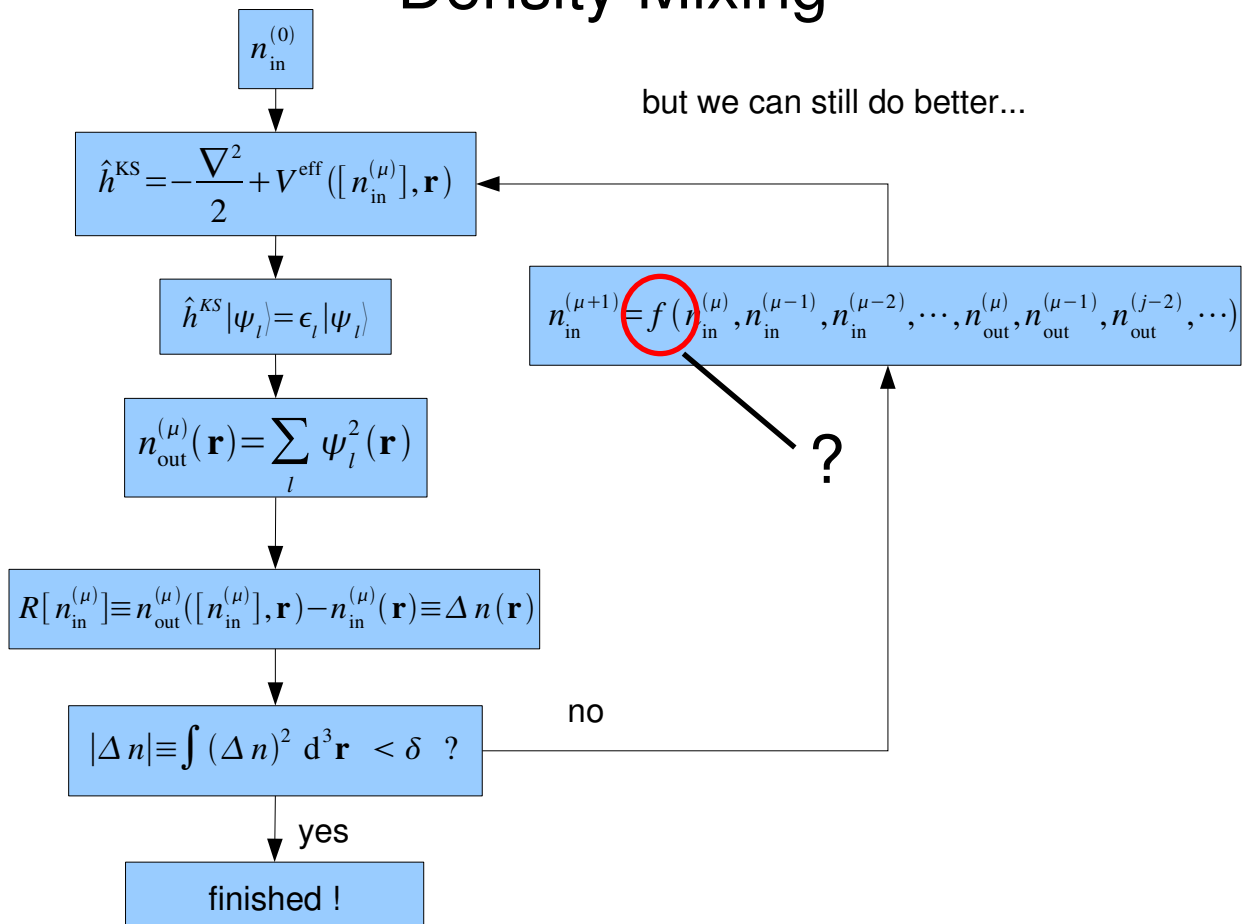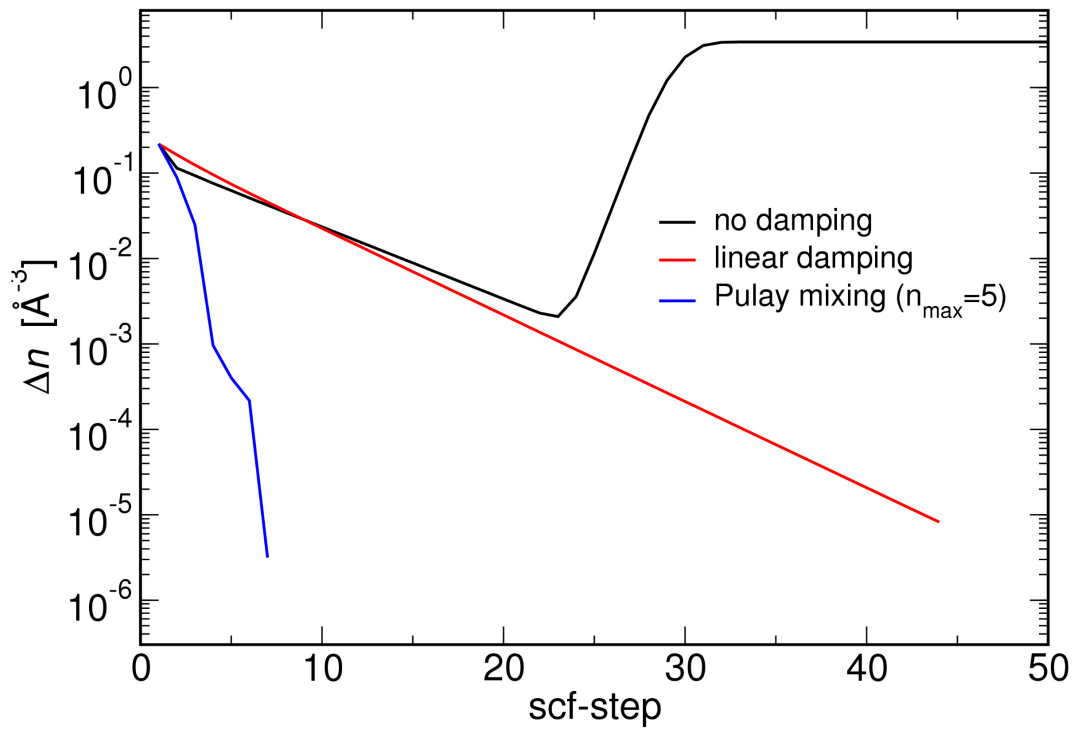
but we can still do better...

$$n_{\text{in}}^{(0)}$$

$$\hat{h}^{\text{KS}} = -\frac{\nabla^2}{2} + V^{\text{eff}}([\,n_{\text{in}}^{(\mu)}], \mathbf{r})$$

$$n_{\text{in}}^{(\mu+1)} = f\left(n_{\text{in}}^{(\mu)}, n_{\text{in}}^{(\mu-1)}, n_{\text{in}}^{(\mu-2)}, \cdots, n_{\text{out}}^{(\mu)}, n_{\text{out}}^{(\mu-1)}, n_{\text{out}}^{(j-2)}, \cdots\right)$$

$$\hat{h}^{KS}|\psi_l\rangle = \epsilon_l|\psi_l\rangle$$

$$n_{\text{out}}^{(\mu)}(\mathbf{r}) = \sum_l \psi_l^2(\mathbf{r})$$

$$R[\,n_{\text{in}}^{(\mu)}] \equiv n_{\text{out}}^{(\mu)}([\,n_{\text{in}}^{(\mu)}], \mathbf{r}) - n_{\text{in}}^{(\mu)}(\mathbf{r}) \equiv \Delta n(\mathbf{r})$$

$$|\Delta n| \equiv \int (\Delta n)^2 \, \mathrm{d}^3\mathbf{r} \; < \delta \; ?$$

no

yes

finished !

Density Mixing

but we can still do better...

$$n_{\text{in}}^{(0)}$$

$$\hat{h}^{\text{KS}} = -\frac{\nabla^2}{2} + V^{\text{eff}}([\,n_{\text{in}}^{(\mu)}], \mathbf{r})$$

$$n_{\text{in}}^{(\mu+1)} = f\left(n_{\text{in}}^{(\mu)}, n_{\text{in}}^{(\mu-1)}, n_{\text{in}}^{(\mu-2)}, \cdots, n_{\text{out}}^{(\mu)}, n_{\text{out}}^{(\mu-1)}, n_{\text{out}}^{(j-2)}, \cdots\right)$$

?

$$\hat{h}^{KS}|\psi_l\rangle = \epsilon_l|\psi_l\rangle$$

$$n_{\text{out}}^{(\mu)}(\mathbf{r}) = \sum_l \psi_l^2(\mathbf{r})$$

$$R[\,n_{\text{in}}^{(\mu)}] \equiv n_{\text{out}}^{(\mu)}([\,n_{\text{in}}^{(\mu)}], \mathbf{r}) - n_{\text{in}}^{(\mu)}(\mathbf{r}) \equiv \Delta n(\mathbf{r})$$

$$|\Delta n| \equiv \int (\Delta n)^2 \, \mathrm{d}^3\mathbf{r} \; < \delta \; ?$$

no

yes

finished !

Density Mixing

$N_2 = 1.1$ Å PBE

The Pulay Mixer

- take more previous densities into account under the constraint of norm conservation

$$n_{in}^{opt} = \sum_\mu \alpha_\mu \, n_{in}^{(\mu)} \qquad\qquad \sum_\mu \alpha_\mu = 1$$

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

The Pulay Mixer

- take more previous densities into account under the constraint of norm conservation

$$n_{in}^{opt} = \sum_{\mu} \alpha_{\mu} n_{in}^{(\mu)} \qquad\qquad \sum_{\mu} \alpha_{\mu} = 1$$

- main assumption: the charge density residual is linear w.r.t. to the density

$$R[n_{in}^{opt}] = R\left[\sum_{\mu} \alpha_{\mu} n_{in}^{(\mu)}\right] = \sum_{\mu} \alpha_{\mu} R\left[n_{in}^{(\mu)}\right]$$

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

The Pulay Mixer

- take more previous densities into account under the constraint of norm conservation

$$n_{in}^{opt} = \sum_{\mu} \alpha_{\mu} n_{in}^{(\mu)} \qquad\qquad \sum_{\mu} \alpha_{\mu} = 1$$

- main assumption: the charge density residual is linear w.r.t. to the density

$$R[n_{in}^{opt}] = R\left[\sum_{\mu} \alpha_{\mu} n_{in}^{(\mu)}\right] = \sum_{\mu} \alpha_{\mu} R\left[n_{in}^{(\mu)}\right]$$

- coefficients α are determined by minimizing the residual

$$\frac{\partial}{\partial \alpha_l}\left(\left\langle R\left[n_{in}^{opt}\right] \middle| R\left[n_{in}^{opt}\right]\right\rangle - \lambda \sum_{\mu} \alpha_{\mu}\right) = 0$$

system of equations for α

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

The Pulay Mixer

- illustration in one dimension with two densities

$$R_1 \equiv R\left[n_{in}^{(1)}\right]$$

$$R_2 \equiv R\left[n_{in}^{(2)}\right]$$

$$\left\langle R\left[n_{in}^{opt}\right]\middle| R\left[n_{in}^{opt}\right]\right\rangle = (\alpha R_1 + (1-\alpha) R_2)^2$$

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

The Pulay Mixer

- illustration in one dimension with two densities

$$R_1 \equiv R\left[n_{in}^{(1)}\right]$$

$$R_2 \equiv R\left[n_{in}^{(2)}\right]$$

$$\left\langle R\left[n_{in}^{opt}\right]\middle| R\left[n_{in}^{opt}\right]\right\rangle = (\alpha R_1 + (1-\alpha) R_2)^2$$
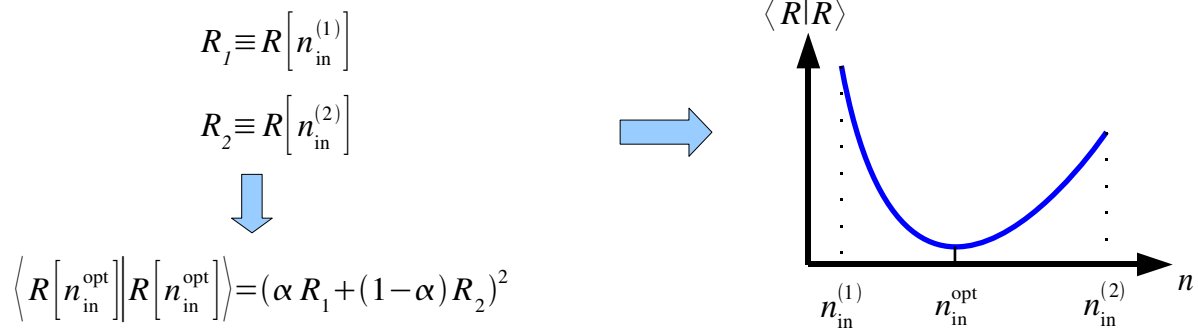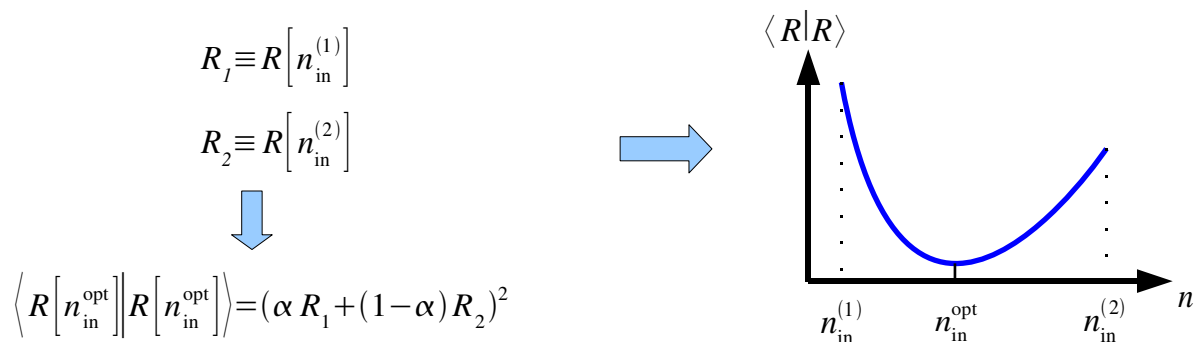
- to prevent trapping in subspace, add fraction of residual

$$n_{in}^{(\mu+1)} = n_{in}^{opt} + \gamma R\left[n_{in}^{opt}\right]$$

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

The Pulay Mixer

- additional preconditioner might accelerate convergence

$$n_{in}^{(\mu+1)} = n_{in}^{opt} + \hat{G}\,R\left[n_{in}^{opt}\right]$$

preconditioning matrix: $\quad \hat{G} = A\,\dfrac{k^2}{k^2 + k_0^2}$

⟹ long-range components of density changes are stronger damped than short-range

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

The Pulay Mixer

- additional preconditioner might accelerate convergence

$$n_{in}^{(\mu+1)} = n_{in}^{opt} + \hat{G}\,R\left[n_{in}^{opt}\right]$$

preconditioning matrix: $\quad \hat{G} = A\,\dfrac{k^2}{k^2 + k_0^2}$

⟹ long-range components of density changes are stronger damped than short-range

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

- two spin channels are to be mixed in the case of spin polarization

charge density: $\quad n^{\uparrow}(\boldsymbol{r}) + n^{\downarrow}(\boldsymbol{r})$ $\qquad\qquad$ spin density: $\quad n^{\uparrow}(\boldsymbol{r}) - n^{\downarrow}(\boldsymbol{r})$

⬇

combine both channels to <u>one</u> matrix !

$$\left\langle R\left[n^{\uparrow},n^{\downarrow}\right]\middle|R\left[n^{\uparrow},n^{\downarrow}\right]\right\rangle = \sum_{\sigma} \int \mathrm{d}\mathbf{r}\,\left(\Delta n^{\sigma}\right)^2(\mathbf{r}) \qquad\qquad \Delta n^{\sigma}(\mathbf{r}) = n_{out}^{\sigma}\left(\left[n_{in}^{\sigma}\right],(\mathbf{r})\right) - n_{in}^{\sigma}(\mathbf{r})$$

⟹ stable convergence

The Broyden Mixer

secant method

$$f(x_n) - f(x_{n-1}) \approx f'(x_n)(x_n - x_{n-1})$$

$$f(x_{n+1}) \overset{!}{=} 0$$

$$\Rightarrow \quad x_{n+1} = x_n - \frac{1}{f'(x_n)} f(x_n)$$

The Broyden Mixer

secant method
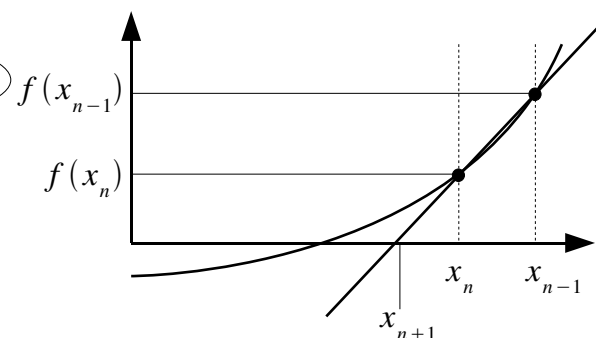
$$f(x_n) - f(x_{n-1}) \approx f'(x_n)(x_n - x_{n-1})$$

$$f(x_{n+1}) \overset{!}{=} 0$$

$$\Rightarrow \quad x_{n+1} = x_n - \frac{1}{f'(x_n)} f(x_n)$$

$$n_{in}^{(\mu+1)} = n_{in}^{(\mu)} + \mathbf{G}^{(\mu)} R[n_{in}^{(\mu)}]$$

approximate Jacobian: $\quad \Delta n_{in}^{(\mu)} + \mathbf{G}^{(\mu)} \Delta R[n_{in}^{(\mu)}] = 0$

D.D. Johnson, Phys. Rev. B **38** (1988), 12807

The Broyden Mixer

secant method
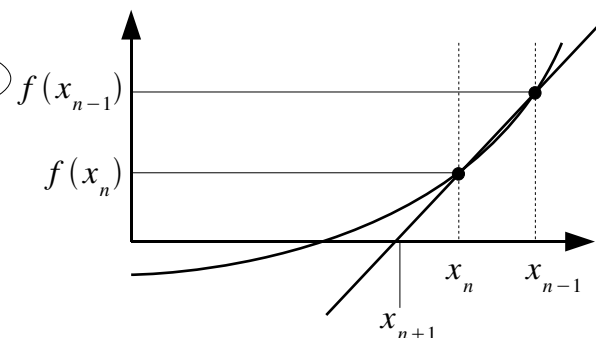
$$f(x_n) - f(x_{n-1}) \approx f'(x_n)(x_n - x_{n-1})$$

$$f(x_{n+1}) \overset{!}{=} 0$$

$$x_{n+1} = x_n - \frac{1}{f'(x_n)} f(x_n)$$

$$n_{\text{in}}^{(\mu+1)} = n_{\text{in}}^{(\mu)} + \mathbf{G}^{(\mu)} R[n_{\text{in}}^{(\mu)}]$$

approximate Jacobian: $\Delta n_{\text{in}}^{(\mu)} + \mathbf{G}^{(\mu)} \Delta R[n_{\text{in}}^{(\mu)}] = 0$

under-determined

$$\frac{\partial}{\partial G_{ij}} \|\mathbf{G}^{(\mu+1)} - \mathbf{G}^{(\mu)}\| = 0$$

"as little change as possible"

$$\mathbf{G}^{(\mu+1)} = \mathbf{G}^{(\mu)} - \left(|\Delta n^{(\mu)}\rangle + \mathbf{G}^{(\mu)} |\Delta R^{(\mu)}\rangle \right) \langle \Delta R^{(\mu)}|$$

D.D. Johnson, Phys. Rev. B **38** (1988), 12807

The Broyden Mixer

secant method

$$f(x_n) - f(x_{n-1}) \approx f'(x_n)(x_n - x_{n-1})$$

$$f(x_{n+1}) \overset{!}{=} 0$$

$$x_{n+1} = x_n - \frac{1}{f'(x_n)} f(x_n)$$

$$n_{\text{in}}^{(\mu+1)} = n_{\text{in}}^{(\mu)} + \mathbf{G}^{(\mu)} R[n_{\text{in}}^{(\mu)}]$$

approximate Jacobian: $\Delta n_{\text{in}}^{(\mu)} + \mathbf{G}^{(\mu)} \Delta R[n_{\text{in}}^{(\mu)}] = 0$

under-determined

$$\frac{\partial}{\partial G_{ij}} \|\mathbf{G}^{(\mu+1)} - \mathbf{G}^{(\mu)}\| = 0$$

"as little change as possible"

$$\mathbf{G}^{(\mu+1)} = \mathbf{G}^{(\mu)} - \left(|\Delta n^{(\mu)}\rangle + \mathbf{G}^{(\mu)} |\Delta R^{(\mu)}\rangle \right) \langle \Delta R^{(\mu)}|$$

modified Broyden: minimize $E = w_0 \left\| \mathbf{G}^{(\mu+1)} - \mathbf{G}^{(\mu)} \right\|^2 + \sum_i w_i \left\| \Delta n^{(i)} + \mathbf{G}^{(\mu+1)} \Delta R^{(i)} \right\|^2$
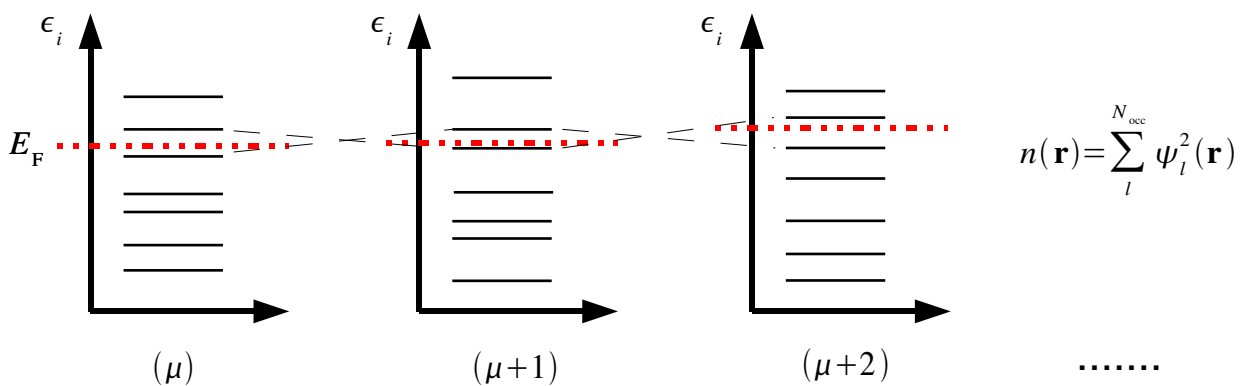
D.D. Johnson, Phys. Rev. B **38** (1988), 12807

Electronic Smearing
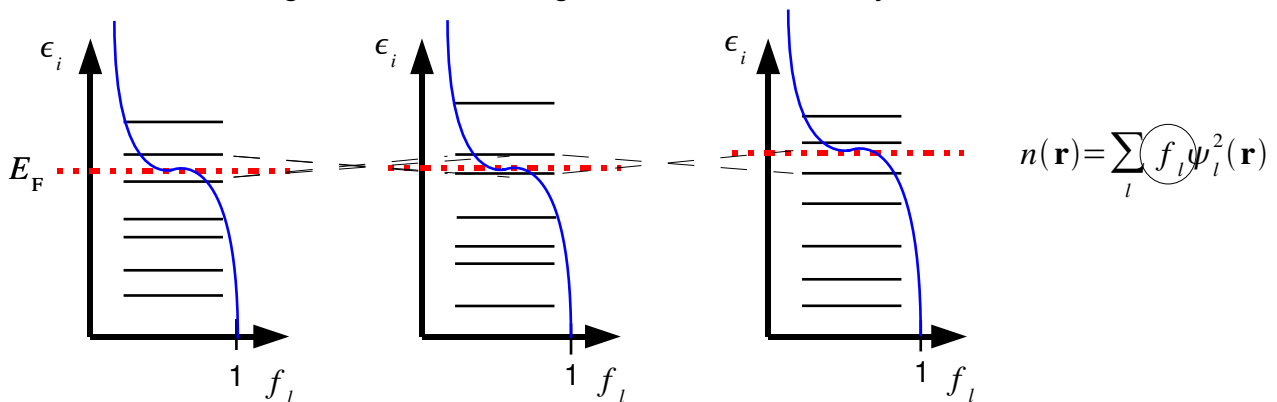
- level crossing may lead to charge sloshing

$$n(\mathbf{r})=\sum_{l}^{N_{occ}} \psi_l^2(\mathbf{r})$$

(μ) $(\mu+1)$ $(\mu+2)$ ....

Electronic Smearing

- level crossing may lead to charge sloshing

$$n(\mathbf{r})=\sum_{l}^{N_{occ}} \psi_l^2(\mathbf{r})$$

(μ) $(\mu+1)$ $(\mu+2)$ ....
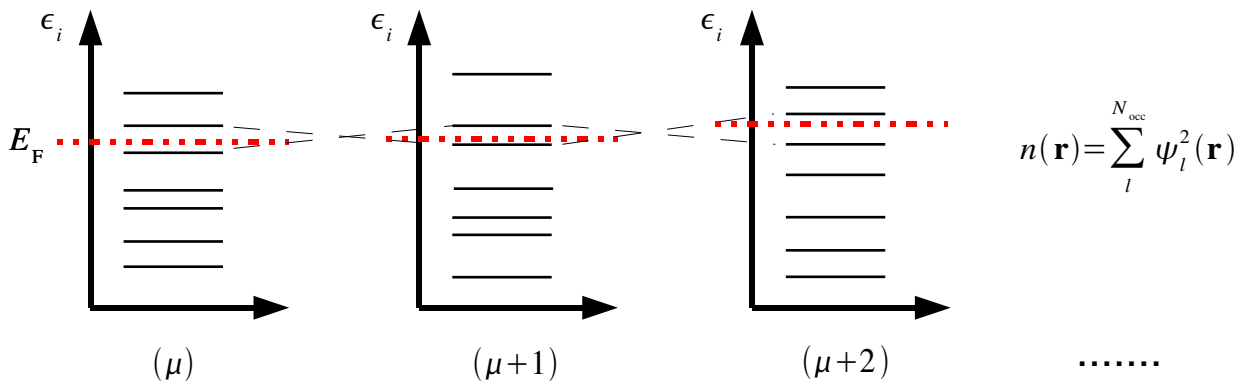
- electronic smearing softens the change of electronic density

$$n(\mathbf{r})=\sum_{l} f_l \psi_l^2(\mathbf{r})$$

$1 \; f_l$ $1 \; f_l$ $1 \; f_l$

Electronic Smearing

Fermi-Dirac[1]:	$f\left(\dfrac{\epsilon - E_F}{\sigma}\right) = \dfrac{1}{\exp\left((\epsilon - E_F)/\sigma\right) + 1}$

Gaussian[2]:	$f\left(\dfrac{\epsilon - E_F}{\sigma}\right) = \dfrac{1}{2}\left(1 - \mathrm{erf}\left[\dfrac{\epsilon - E_F}{\sigma}\right]\right)$

Methfessel-Paxton[3]: $\quad f_0(x) = \dfrac{1}{2}\left(1 - \mathrm{erf}(x)\right)$ $\qquad\qquad x = \dfrac{\epsilon - E_F}{\sigma}$

$$f_N(x) = f_0(x) + \sum_{m=1}^{N} A_m H_{2m-1}(x)\, e^{x^2} \xrightarrow[N \to \infty]{} \Theta(x)$$

[1] N. Mermin, Phys. Rev. **137** (1965), A1441

[2] C.-L. Fu, K.-H. Ho, Phys. Rev. B **28** (1983), 5480

[3] M. Methfessel, A. Paxton, Phys. Rev. B **40** (1989), 3616

Electronic Smearing

• Kohn-Sham functional not a variational quantity anymore

$$\tilde{E}[\{c\}] \quad \Longrightarrow \quad \tilde{E}[\{c\}, \{f\}] \quad \Longrightarrow \quad \left.\frac{\partial \tilde{E}}{\partial f_i}\right|_{n_0} \neq 0$$

Electronic Smearing

- Kohn-Sham functional not a variational quantity anymore

$$\tilde{E}[\{c\}] \implies \tilde{E}[\{c\},\{f\}] \implies \left.\frac{\partial \tilde{E}}{\partial f_i}\right|_{n_0} \neq 0$$

- electronic *free* energy F is variational

$$F = E - \underbrace{\sigma S(\sigma,\{f_n\})}_{\text{entropy term}} \implies \left.\frac{\partial F}{\partial f_i}\right|_{n_0} = 0$$
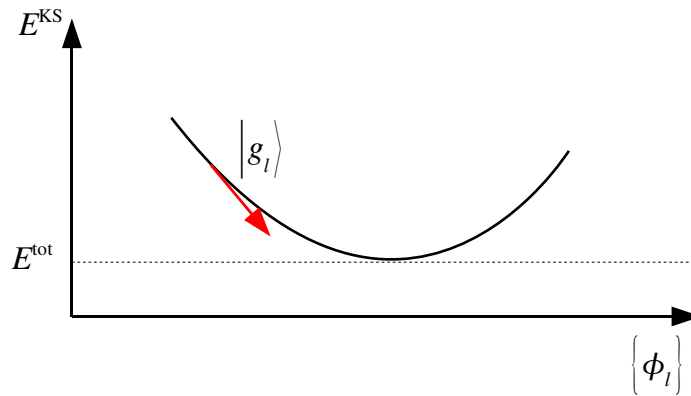

Electronic Smearing

- Kohn-Sham functional not a variational quantity anymore

$$\tilde{E}[\{c\}] \implies \tilde{E}[\{c\},\{f\}] \implies \left.\frac{\partial \tilde{E}}{\partial f_i}\right|_{n_0} \neq 0$$

- electronic *free* energy F is variational

$$F = E - \underbrace{\sigma S(\sigma,\{f_n\})}_{\text{entropy term}} \implies \left.\frac{\partial F}{\partial f_i}\right|_{n_0} = 0$$

- systems with <u>continuous density of states</u> at the Fermi edge:

$$F(\sigma) = E_{\sigma=0} + \begin{cases} O(\sigma^2) \\ O(\sigma^{2+N}) \end{cases} \qquad E_{\sigma=0} \approx \begin{cases} \dfrac{1}{2}(F(\sigma)+E(\sigma)) & \text{Fermi-Dirac / Gaussian} \\[2ex] \dfrac{1}{N+2}((N+1)F(\sigma)+E(\sigma)) & \text{Methfessel-Paxton} \end{cases}$$

$$\implies \boxed{\text{desired total energy } E \text{ for zero smearing can be backextrapolated}}$$

M. Gillan, J. Phys.: Condens. Matter **1** (1989), 689

M. Methfessel, A. Paxton, Phys. Rev. B **40** (1989), 3616

F. Wagner, T. Laloyaux, M. Scheffler, Phys. Rev. B **57** (1998), 2102

Alternative: Direct Minimization

- constrained minimum of Kohn-Sham functional:

$$\tilde{E} = E^{\text{KS}} - \sum_l \epsilon_l \left(\langle \phi_l | \phi_l \rangle - 1 \right) - E_{\text{F}} \left(\sum_l f_l - N \right)$$

- gradient:

$$\frac{\delta \tilde{E}}{\delta \langle \phi_l |} \equiv | g_l \rangle = f_l \left(\hat{h}^{\text{KS}} - \epsilon_l \right) | \phi_l \rangle$$
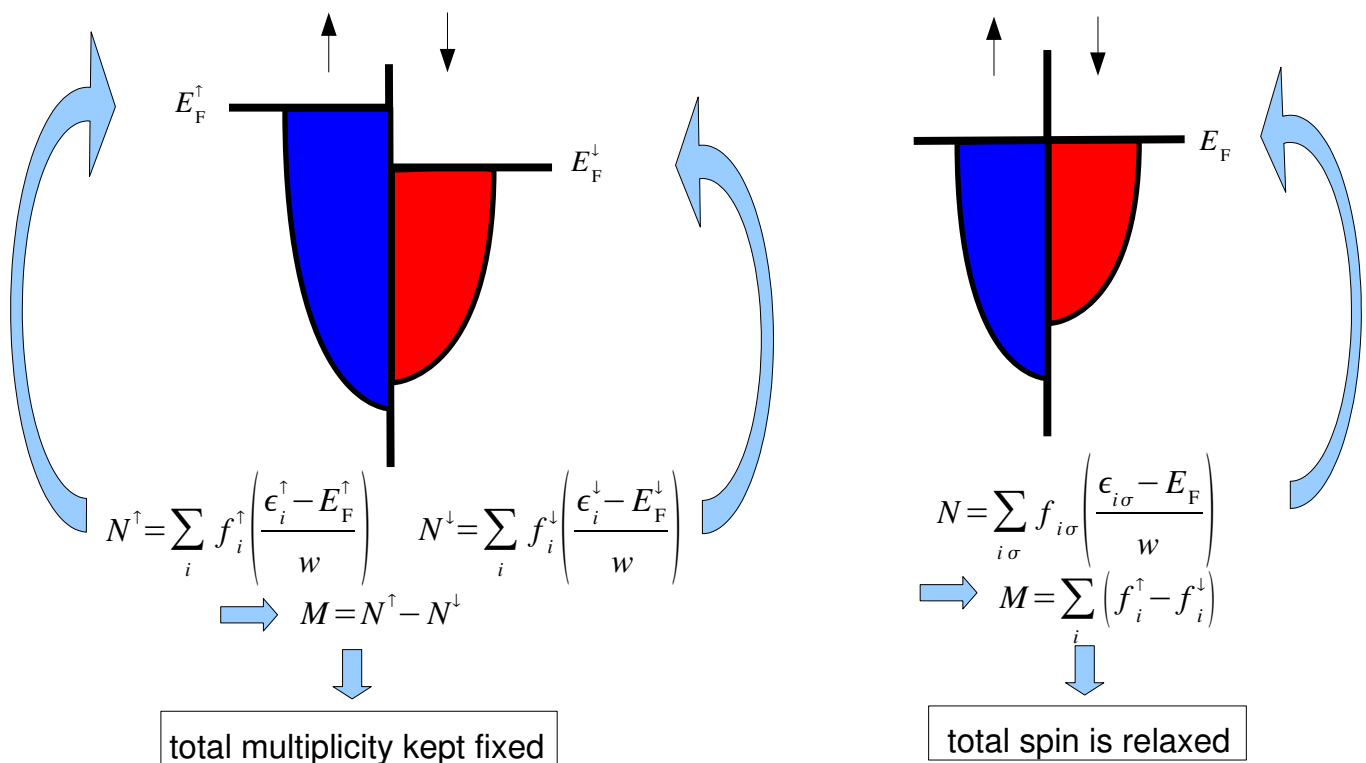
- steepest-descent:

$$\left| \phi_l^{(\mu+1)} \right\rangle = \left| \phi_l^{(\mu)} \right\rangle - \Delta \left| g_l \right\rangle$$

Fixed Spin vs. Unconstrained Spin

$$\left[-\frac{1}{2} \nabla^2 + V_{\text{eff}}^{\sigma}(\mathbf{r}) \right] \phi_{i\sigma} = \epsilon_{i\sigma} \phi_{i\sigma} \qquad V_{\text{XC}}^{\sigma}(\mathbf{r}) \equiv \frac{\partial E_{\text{XC}}}{\partial n^{\sigma}(\mathbf{r})}$$

$$N^{\uparrow} = \sum_i f_i^{\uparrow} \left(\frac{\epsilon_i^{\uparrow} - E_{\text{F}}^{\uparrow}}{w} \right) \qquad N^{\downarrow} = \sum_i f_i^{\downarrow} \left(\frac{\epsilon_i^{\downarrow} - E_{\text{F}}^{\downarrow}}{w} \right)$$

$$\Rightarrow M = N^{\uparrow} - N^{\downarrow}$$

total multiplicity kept fixed

$$N = \sum_{i\sigma} f_{i\sigma} \left(\frac{\epsilon_{i\sigma} - E_{\text{F}}}{w} \right)$$

$$\Rightarrow M = \sum_i \left(f_i^{\uparrow} - f_i^{\downarrow} \right)$$

total spin is relaxed

Fixed Spin vs. Unconstrained Spin
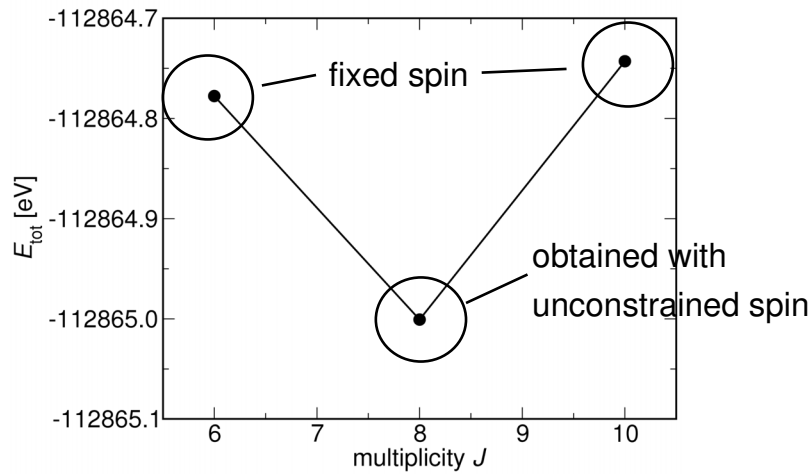
Co_3 PBE
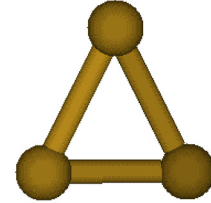
⟹ obtained multiplicity the lowest one

Atomic Forces

- total energy of the system is the total minimum of the Kohn-Sham-functional under the normalization constraint:

$$E_{tot} = \min_{\{c\}}\left(E^{KS} - \sum_{i\sigma} \epsilon_{i\sigma}\left(\langle \phi_{i\sigma} | \phi_{i\sigma} \rangle - 1 \right) \right) = \tilde{E}\left[\{c_0\{\mathbf{R}_\alpha\}\}, \{\epsilon\{\mathbf{R}_\alpha\}\}, \{\mathbf{R}_\alpha\} \right]$$

$$\left. \frac{\partial \tilde{E}}{\partial \{c\}} \right|_{c=c_0} = 0 \qquad \left. \frac{\partial \tilde{E}}{\partial \{\epsilon\}} \right|_{c=c_0} = 0$$

Atomic Forces

- total energy of the system is the total minimum of the Kohn-Sham-functional under the normalization constraint:

$$E_{\text{tot}} = \min_{\{c\}} \left(E^{\text{KS}} - \sum_{i\sigma} \epsilon_{i\sigma} \left(\langle \phi_{i\sigma} | \phi_{i\sigma} \rangle - 1 \right) \right) = \tilde{E} \left[\left\{ c_0 \left[\mathbf{R}_\alpha \right] \right\}, \left\{ \epsilon \left[\mathbf{R}_\alpha \right] \right\}, \left\{ \mathbf{R}_\alpha \right\} \right] \qquad \left. \frac{\partial \tilde{E}}{\partial \{c\}} \right|_{c=c_0} = 0 \qquad \left. \frac{\partial \tilde{E}}{\partial \{\epsilon\}} \right|_{c=c_0} = 0$$

- since the Kohn-Sham-functional is variational, only the partial derivative remains:

$$\mathbf{F}_\alpha = -\frac{\mathrm{d}\, E_{\text{tot}}}{\mathrm{d}\, \mathbf{R}_\alpha}$$

$$= -\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha} - \sum_{\{c_0\}} \underbrace{\frac{\partial \tilde{E}}{\partial \{c_0\}}}_{=0} \frac{\partial \{c_0\}}{\partial \mathbf{R}_\alpha} - \sum_{\{\epsilon\}} \underbrace{\frac{\partial \tilde{E}}{\partial \{\epsilon\}}}_{=0} \frac{\partial \{\epsilon\}}{\partial \mathbf{R}_\alpha}$$

$$\boxed{ = -\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha} = -\frac{\partial E^{\text{KS}}}{\partial \mathbf{R}_\alpha} - \frac{\partial}{\partial \mathbf{R}_\alpha} \sum_{i\sigma} \epsilon_{i\sigma} \left(\langle \phi_{i\sigma} | \phi_{i\sigma} \rangle - 1 \right) }$$

The Individual Force Contributions

- classical Coulomb interaction:

$$\mathbf{F}_{\text{HF},\alpha} = Z_\alpha \sum_{\beta \neq \alpha} Z_\beta \frac{\mathbf{R}_\alpha - \mathbf{R}_\beta}{\left| \mathbf{R}_\alpha - \mathbf{R}_\beta \right|^3} - Z_\alpha \int \mathrm{d}\mathbf{r}\, n(\mathbf{r}) \frac{\mathbf{R}_\alpha - \mathbf{r}}{\left| \mathbf{R}_\alpha - \mathbf{r} \right|^3}$$

- Pulay-correction due to incomplete or nuclear-dependent basis set:

$$\mathbf{F}_{\text{Pulay},\alpha} = -2 \sum_{l\sigma} \left\langle \frac{\partial \psi_{l\sigma}}{\partial \mathbf{R}_\alpha} \middle| \hat{h}^{\text{KS}} - \epsilon_{l\sigma} \middle| \psi_{l\sigma} \right\rangle \qquad \Longrightarrow \quad \text{"moving basis functions"}$$

- Multipole-correction due to "incomplete" Hartree-potential:

$$\mathbf{F}_{\text{MP},\alpha} = -\int \mathrm{d}\mathbf{r} \left(n(\mathbf{r}) - n^{\text{MP}}(\mathbf{r}) \right) \frac{\partial V_{\text{H}} \left[n^{\text{MP}} \right]}{\partial \mathbf{R}_\alpha} \qquad \Longrightarrow \quad \text{"moving multipole components"}$$

- gradient-dependency of exchange-correlation energy in case of a GGA functional:

$$\mathbf{F}_{\text{GGA},\alpha} = -\int \mathrm{d}\mathbf{r}\, n(\mathbf{r}) \frac{\partial \epsilon_{\text{XC}}}{\partial \left| \nabla n \right|^2} \frac{\partial \left| \nabla n \right|^2}{\partial \mathbf{R}_\alpha}$$
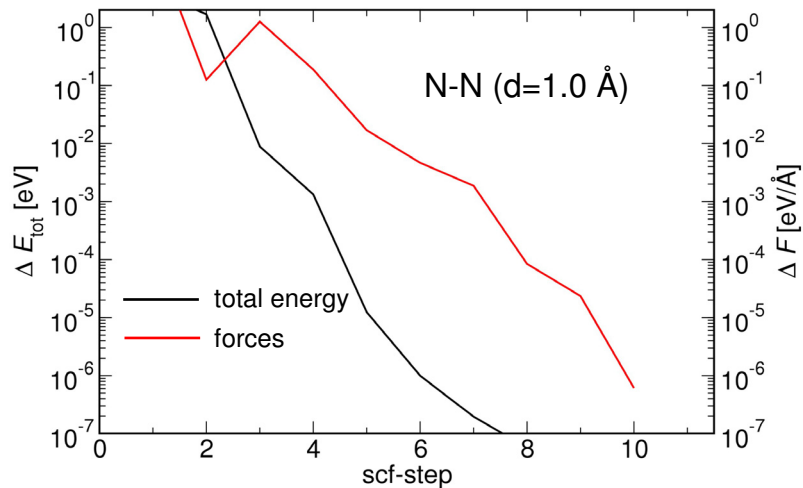
The Convergence of the Forces

- total energy is variational $\Longrightarrow$ error in the density enters total energy to <u>second</u> order

$$\left.\frac{\partial \tilde{E}}{\partial \{c\}}\right|_{c=c_0}=0$$

- atomic forces are <u>not</u> variational $\Longrightarrow$ error in the density enters total energy to <u>first</u> order

$$\left.\frac{\partial \mathbf{F}_\alpha}{\partial \{c\}}\right|_{c=c_0}=\left.\frac{\partial^2 \tilde{E}}{\partial \mathbf{R}_\alpha \partial \{c\}}\right|_{c=c_0}\neq 0$$

$\Longrightarrow$ convergence of atomic forces more expensive than total energy

---

Forces and Electronic Smearing

- Kohn-Sham functional not variational w.r.t. the occupation numbers

$$\tilde{E}[\{c\},\{\epsilon\},\{\mathbf{R}\},\{f\},E_F] \qquad \left.\frac{\partial \tilde{E}}{\partial f_i}\right|_{n_0}\neq 0$$

$\Longrightarrow$ atomic forces are not consistent with the total energy of the system !

$$\mathbf{F}_\alpha=-\frac{d E_{tot}}{d \mathbf{R}_\alpha}$$

$$=-\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha}-\sum_{\{f\}}\frac{\partial \tilde{E}}{\underbrace{\partial \{f\}}_{\neq 0}}\frac{\partial \{f\}}{\partial \mathbf{R}_\alpha}-\underbrace{\frac{\partial \tilde{E}}{\partial E_F}\frac{\partial E_F}{\partial \mathbf{R}_\alpha}}_{=0}\quad \neq -\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha}$$

$\Longrightarrow$ atomic forces are consistent with the electronic <u>free energy</u>

$$\left.\frac{\partial F}{\partial f_i}\right|_{n_0}=0 \qquad\qquad \mathbf{F}_\alpha=-\frac{\partial F}{\partial \mathbf{R}_\alpha}$$

M. Weinert and J.W. Davenport, Phys. Rev. B **45** (1992), 13709

R.M. Wentzcovitch, J.L. Martin and P.B. Allen, Phys. Rev. B **45** (1992), 11372

The Second Derivative and Vibrations

analytical

- first derivative was simple

$$\mathbf{F}_\alpha = -\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha}$$

The Second Derivative and Vibrations

analytical

- first derivative was simple

$$\mathbf{F}_\alpha = -\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha}$$

- second derivative more nasty

$$\mathbf{H}_{\alpha\beta} = \frac{d^2 E_{\text{tot}}}{d\mathbf{R}_\alpha \, d\mathbf{R}_\beta} = -\frac{d\mathbf{F}_\alpha}{d\mathbf{R}_\beta} = \frac{\partial^2 \tilde{E}}{\partial \mathbf{R}_\alpha \partial \mathbf{R}_\beta} + \sum_{\{c\}} \frac{\partial^2 \tilde{E}}{\partial \mathbf{R}_\alpha \partial \{c\}} \frac{\partial \{c_0\}}{\partial \mathbf{R}_\beta}$$

e.g. A. Komornicki and G. Fitzgerald, J. Chem. Phys. **98** (2), 1398 (1993)

The Second Derivative and Vibrations

analytical

- first derivative was simple

$$\mathbf{F}_\alpha = -\frac{\partial \tilde{E}}{\partial \mathbf{R}_\alpha}$$

- second derivative more nasty

$$\mathbf{H}_{\alpha\beta} = \frac{d^2 E_{\text{tot}}}{d\mathbf{R}_\alpha d\mathbf{R}_\beta} = -\frac{d\mathbf{F}_\alpha}{d\mathbf{R}_\beta} = \frac{\partial^2 \tilde{E}}{\partial\mathbf{R}_\alpha \partial\mathbf{R}_\beta} + \sum_{\{c\}} \frac{\partial^2 \tilde{E}}{\partial\mathbf{R}_\alpha \partial\{c\}} \boxed{\frac{\partial\{c_0\}}{\partial\mathbf{R}_\beta}}$$

- set of coupled equations needs to be solved

$$\sum_{\{c\}} \frac{\partial^2 E}{\partial\{c\}^2} \frac{\partial\{c_0\}}{\partial\mathbf{R}_\alpha} = -\frac{\partial^2 E}{\partial\mathbf{R}_\alpha \partial\{c\}}$$

e.g. A. Komornicki and G. Fitzgerald, J. Chem. Phys. **98** (2), 1398 (1993)

The Second Derivative and Vibrations

numerical

$$\frac{d\mathbf{F}_\alpha}{d\mathbf{R}_\beta} \approx \frac{\mathbf{F}_\alpha(\mathbf{R}_{\beta,0}+\Delta) - \mathbf{F}_\alpha(\mathbf{R}_{\beta,0}-\Delta)}{2\Delta}$$

$\mathbf{R}_{\beta,0}$
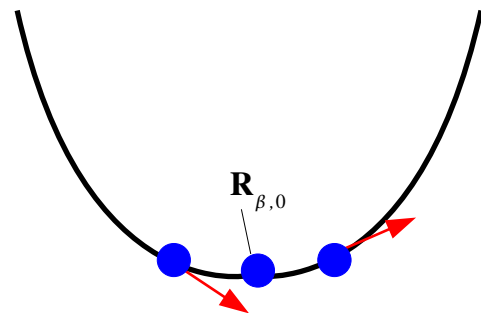
vibrations

$$M_\alpha \ddot{\mathbf{R}}_\alpha = -\frac{d}{d\mathbf{R}_\alpha} E(\{\mathbf{R}\}) \qquad E(\{\mathbf{R}\}) = E(\{\mathbf{R}_0\}) + \frac{1}{2}\sum_{\alpha\beta} \frac{d^2 E}{d\mathbf{R}_\alpha d\mathbf{R}_\beta}\bigg|_{\{\mathbf{R}_0\}} \mathbf{u}_\alpha \mathbf{u}_\beta \qquad \mathbf{u}_\alpha = \mathbf{R}_\alpha - \mathbf{R}_{\alpha,0}$$

$$\mathbf{u}_\alpha(t) = \mathbf{u}_\alpha e^{i\omega t} \qquad \Longrightarrow \qquad \det\left(\mathbf{H}_{\alpha\beta} - \omega^2 M_\alpha\right) \overset{!}{=} 0$$

⟹ talk by Gert von Helden, Friday, 26th June, 11:30,
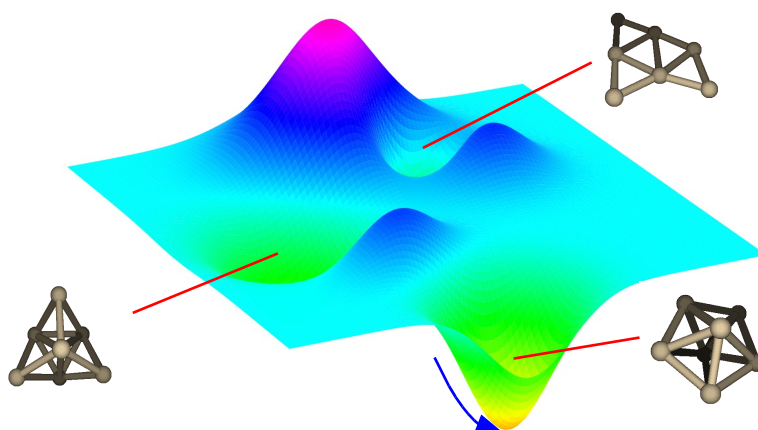"Fingerprinting molecules: Vibrational Spectroscopy, Experiment and Theory"

Local Atomic Structure Optimization

Potential Energy Surface (PES) $E_{tot}\left(\left\{\boldsymbol{R}_\alpha\right\}\right)$
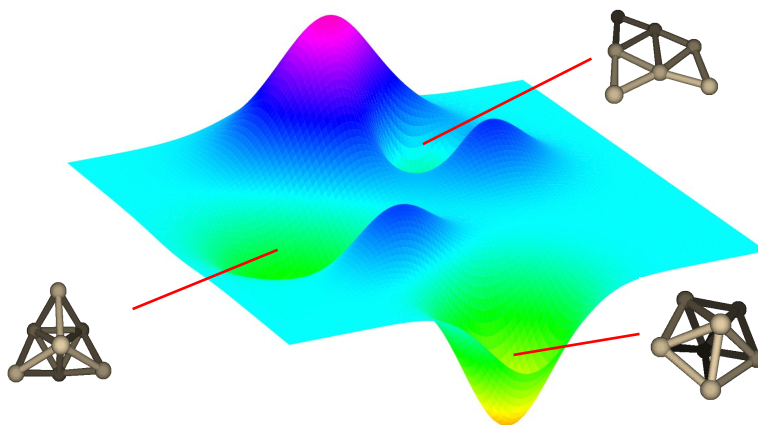
Local Atomic Structure Optimization

Potential Energy Surface (PES) $E_{tot}\left(\left\{\boldsymbol{R}_\alpha\right\}\right)$
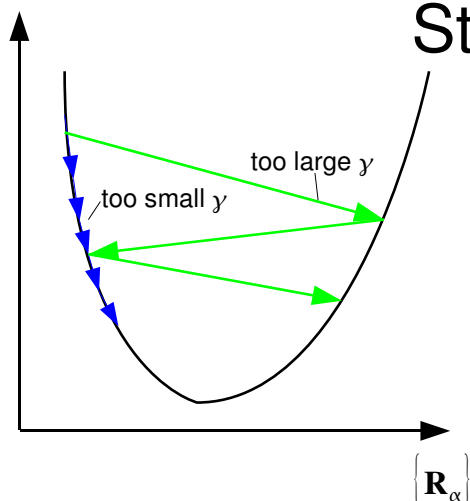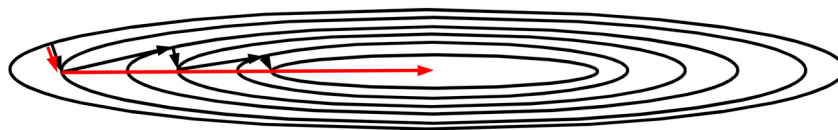
Steepest Descent

$$\mathbf{R}_{\alpha, i+1} = \mathbf{R}_{\alpha, i} + \gamma_i \, \mathbf{F}_\alpha\left(\left\{\mathbf{R}_{\alpha, i}\right\}\right)$$
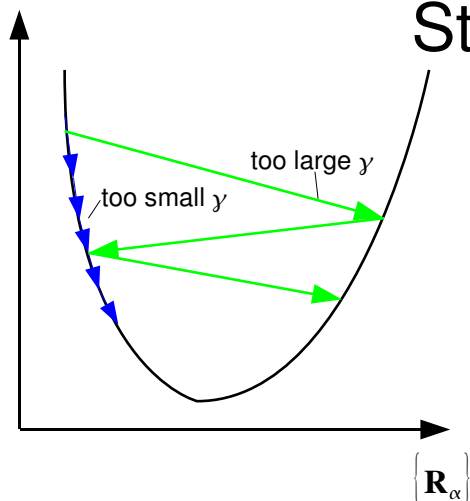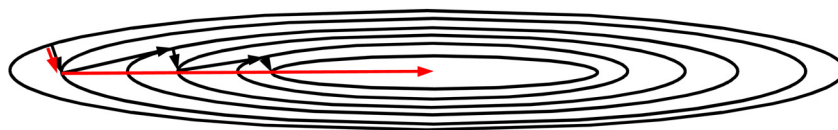

Steepest Descent

too large γ

too small γ

$$\mathbf{R}_{\alpha, i+1} = \mathbf{R}_{\alpha, i} + \gamma_i \, \mathbf{F}_\alpha\left(\left\{\mathbf{R}_{\alpha, i}\right\}\right)$$

$\left\{\mathbf{R}_\alpha\right\}$

Steepest Descent

too large γ

too small γ

$$\mathbf{R}_{\alpha,i+1} = \mathbf{R}_{\alpha,i} + \gamma_i \mathbf{F}_\alpha\left(\left\{\mathbf{R}_{\alpha,i}\right\}\right)$$

$\left\{\mathbf{R}_\alpha\right\}$

ill-conditioned PES might lead to zig-zag behaviour

Steepest Descent

too large γ

too small γ

$$\mathbf{R}_{\alpha,i+1} = \mathbf{R}_{\alpha,i} + \gamma_i \mathbf{F}_\alpha\left(\left\{\mathbf{R}_{\alpha,i}\right\}\right)$$

$\left\{\mathbf{R}_\alpha\right\}$

ill-conditioned PES might lead to zig-zag behaviour

Conjugate Gradient

- take information of previous search directions into account

$$\mathbf{R}_{\alpha,i+1} = \mathbf{R}_{\alpha,i} + \gamma_i \mathbf{G}_{\alpha,i}\left(\left\{\mathbf{R}_{\alpha,i}\right\}\right)$$

line minimization

$$\mathbf{G}_{\alpha,i} = \mathbf{F}_{\alpha,i} + \beta_i \mathbf{G}_{\alpha,i-1}$$

$$\beta_i^{\text{Fletcher-Reeves}} = \frac{\mathbf{F}_{\alpha,i} \circ \mathbf{F}_{\alpha,i}}{\mathbf{F}_{\alpha,i-1} \circ \mathbf{F}_{\alpha,i-1}}$$

$$\beta_i^{\text{Polak-Ribiere}} = \frac{\mathbf{F}_{\alpha,i} \circ \left(\mathbf{F}_{\alpha,i} - \mathbf{F}_{\alpha,i-1}\right)}{\mathbf{F}_{\alpha,i-1} \circ \mathbf{F}_{\alpha,i-1}}$$

- constructions of search directions based upon the assumption that PES is perfectly harmonic

$$\vdots$$

The Broyden-Fletcher-Goldfarb-Shanno Method
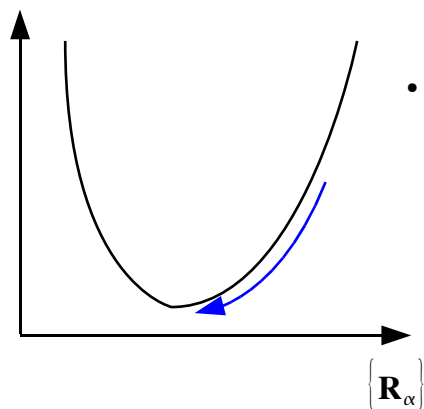
<u>Newton-Method:</u>

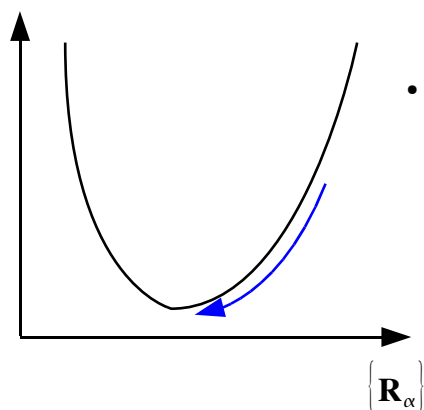• additional to the atomic forces, the Hessian is taken into account:

$$f(\mathbf{x}) = f(\mathbf{x}_i) + (\mathbf{x} - \mathbf{x}_i) \circ \nabla f(\mathbf{x}_i) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_i) \circ \mathbf{A} \circ (\mathbf{x} - \mathbf{x}_i) + \cdots$$

$$\nabla f(\mathbf{x}) = \nabla f(\mathbf{x}_i) + \mathbf{A} \circ (\mathbf{x} - \mathbf{x}_i) + \cdots$$

$$\nabla f(\mathbf{x}) \overset{!}{=} 0 \quad \Longrightarrow \quad \mathbf{x} - \mathbf{x}_i = -\mathbf{A^{-1}} \circ \nabla f(\mathbf{x}_i)$$

$$\left[\mathbf{R}_\alpha\right]$$

W. H. Press et al., Numerical Recipes, 3rd Edition, Cambridge University Press

---

The Broyden-Fletcher-Goldfarb-Shanno Method

<u>Newton-Method:</u>

• additional to the atomic forces, the Hessian is taken into account:

$$f(\mathbf{x}) = f(\mathbf{x}_i) + (\mathbf{x} - \mathbf{x}_i) \circ \nabla f(\mathbf{x}_i) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_i) \circ \mathbf{A} \circ (\mathbf{x} - \mathbf{x}_i) + \cdots$$

$$\nabla f(\mathbf{x}) = \nabla f(\mathbf{x}_i) + \mathbf{A} \circ (\mathbf{x} - \mathbf{x}_i) + \cdots$$

$$\nabla f(\mathbf{x}) \overset{!}{=} 0 \quad \Longrightarrow \quad \mathbf{x} - \mathbf{x}_i = -\mathbf{A^{-1}} \circ \nabla f(\mathbf{x}_i)$$

$$\left[\mathbf{R}_\alpha\right]$$

• "Quasi"-Newton-Scheme: inverse Hessian is approximated iteratively

$$\mathbf{R}_{\alpha,i+1} = \mathbf{R}_{\alpha,i} + \gamma_i \mathbf{G}_{\alpha,i}\left(\left[\mathbf{R}_{\alpha,i}\right]\right) \qquad\qquad \mathbf{G}_{\alpha,i} = \mathbf{H}_i \mathbf{F}_{\alpha,i}$$

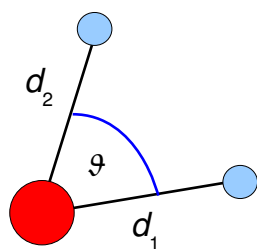$$\mathbf{H}_0 = 1 \qquad \mathbf{R}_{\alpha,i+1} - \mathbf{R}_{\alpha,i} = \mathbf{H}_{i+1} \circ \left(\mathbf{F}_{\alpha,i} - \mathbf{F}_{\alpha,i+1}\right) \qquad \Longrightarrow \qquad \mathbf{H}_{i+1} = \mathbf{H}_i + \cdots$$

• 1 ⇒ perfectly harmonic PES is assumed

• line minimization

W. H. Press et al., Numerical Recipes, 3rd Edition, Cambridge University Press

Cartesian vs. Internal Coordinates

$$\left\{ \mathbf{R}_1 , \mathbf{R}_2 , \mathbf{R}_3 \right\} \rightarrow \left\{ q_1 , q_2 , q_3 \right\} \qquad q_1 = d_1 \quad q_2 = d_2 \quad q_3 = \vartheta$$

$$\mathbf{q} = \mathbf{B} \left(\mathbf{R} - \mathbf{R}_0 \right)$$

$\Longrightarrow \quad \mathbf{R} = \mathbf{R}(\mathbf{q})$ difficult!

(3N-6) x (3N) $\Rightarrow$ not invertable !

P. Pulay, G. Fogarasi, F. Pang, J.E. Boggs, J. Am. Chem. Soc. **101**, 2550 (1979)

Cartesian vs. Internal Coordinates

$$\left\{ \mathbf{R}_1 , \mathbf{R}_2 , \mathbf{R}_3 \right\} \rightarrow \left\{ q_1 , q_2 , q_3 \right\} \qquad q_1 = d_1 \quad q_2 = d_2 \quad q_3 = \vartheta$$

$$\mathbf{q} = \mathbf{B} \left(\mathbf{R} - \mathbf{R}_0 \right)$$

$\Longrightarrow \quad \mathbf{R} = \mathbf{R}(\mathbf{q})$ difficult!

(3N-6) x (3N) $\Rightarrow$ not invertable !
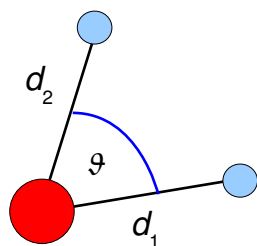
$$E = E_0 - \sum_i \phi_i q_i + \frac{1}{2} \sum_{ij} F^{ij} q_i q_j + \frac{1}{6} \underbrace{\sum_{ijk} F^{ijk} q_i q_j q_k} + \cdots$$

$$\phi_i = -\frac{\partial E}{\partial q_i} \qquad\qquad \text{as "simple" as possible}$$

$$F^{ijk} = -\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k}$$
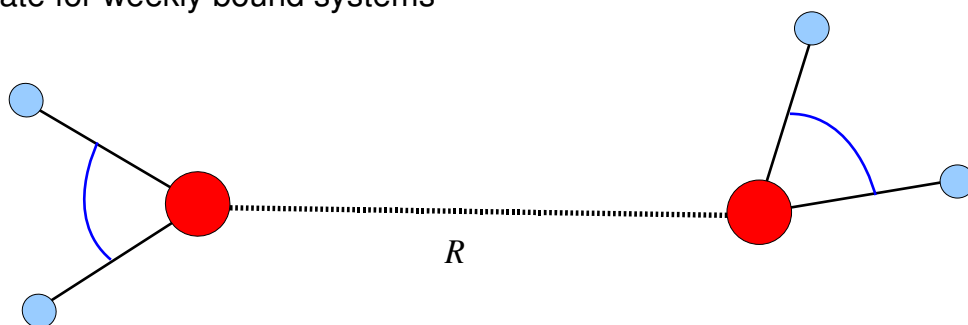
$\Longrightarrow$ faster convergence of the local geometry optimizer

P. Pulay, G. Fogarasi, F. Pang, J.E. Boggs, J. Am. Chem. Soc. **101**, 2550 (1979)

Inverse-Distance Coordinates
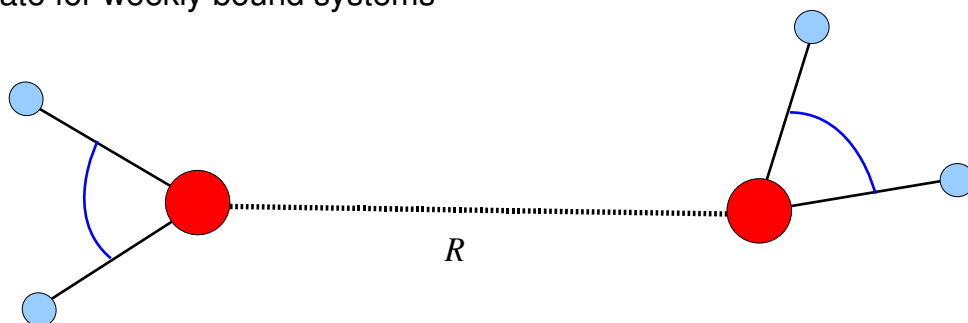
- appropriate for weekly bound systems

P.E. Maslen, J. Chem. Phys. **122**, 014104 (2005)

J. Baker, P. Pulay, J. Chem. Phys. **105** (24), 11100 (1996)

Inverse-Distance Coordinates

- appropriate for weekly bound systems

$$q = \frac{1}{R}$$

$$\frac{\partial E}{\partial q} = -\frac{\partial E}{\partial R} \circ R^2$$
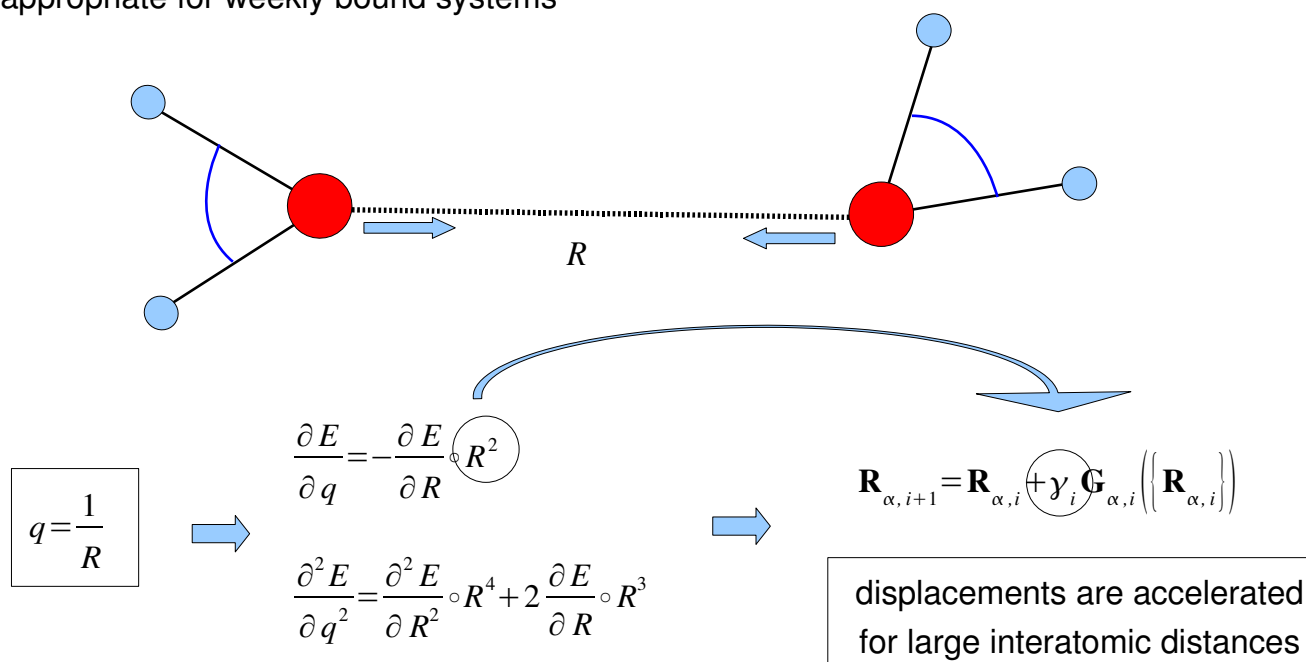
$$\frac{\partial^2 E}{\partial q^2} = \frac{\partial^2 E}{\partial R^2} \circ R^4 + 2 \frac{\partial E}{\partial R} \circ R^3$$

P.E. Maslen, J. Chem. Phys. **122**, 014104 (2005)

J. Baker, P. Pulay, J. Chem. Phys. **105** (24), 11100 (1996)

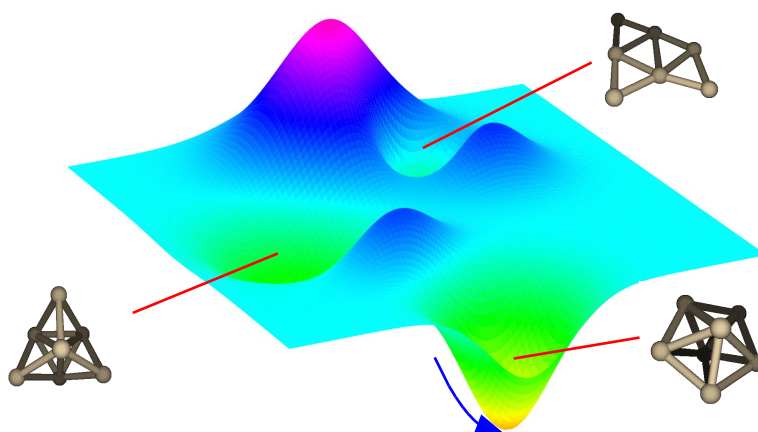# Inverse-Distance Coordinates

- appropriate for weakly bound systems

$$R$$

$$q = \frac{1}{R}$$

$$\frac{\partial E}{\partial q} = -\frac{\partial E}{\partial R} \circ R^2$$

$$\frac{\partial^2 E}{\partial q^2} = \frac{\partial^2 E}{\partial R^2} \circ R^4 + 2\frac{\partial E}{\partial R} \circ R^3$$

$$\mathbf{R}_{\alpha, i+1} = \mathbf{R}_{\alpha, i} + \gamma_i \mathbf{G}_{\alpha, i}\left(\left\{\mathbf{R}_{\alpha, i}\right\}\right)$$

displacements are accelerated for large interatomic distances

P.E. Maslen, J. Chem. Phys. **122**, 014104 (2005)

J. Baker, P. Pulay, J. Chem. Phys. **105** (24), 11100 (1996)

Global Structure Optimization

Potential Energy Surface (PES) $E_{\text{tot}}\left(\left\{\boldsymbol{R}_\alpha\right\}\right)$

Global Structure Optimization

Potential Energy Surface (PES) $\quad E_{\text{tot}}\left(\left\{\boldsymbol{R}_\alpha\right\}\right)$
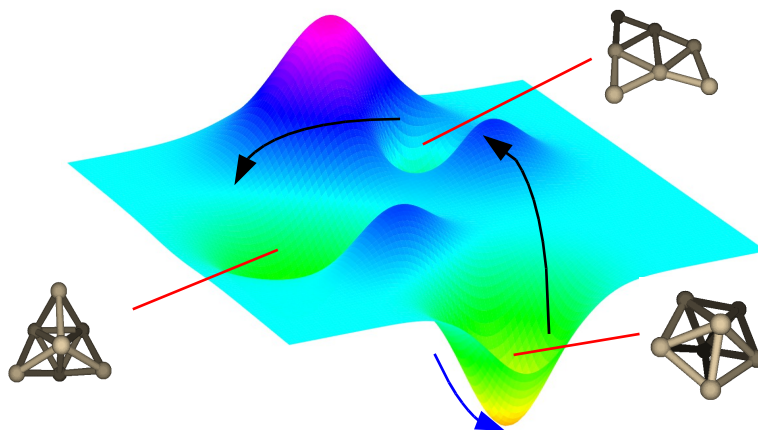
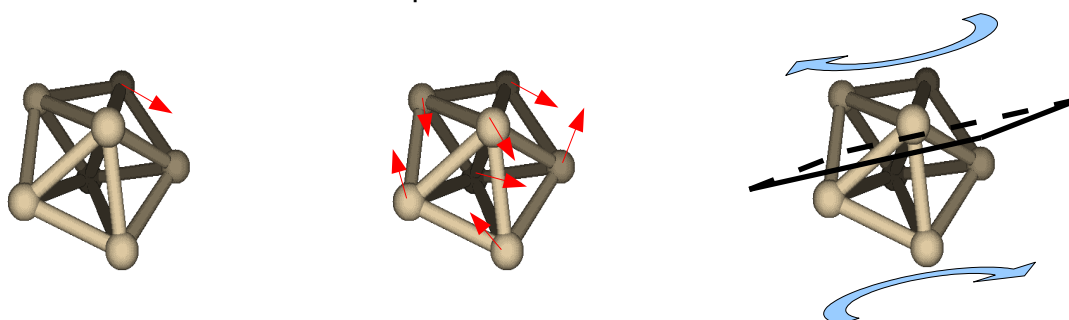
Basin-Hopping

• random variations of atomic coordinates followed by local structural relaxations
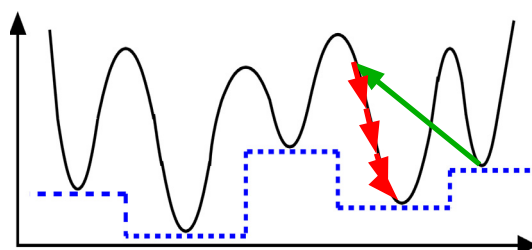
• many different flavours of "trial moves" possible

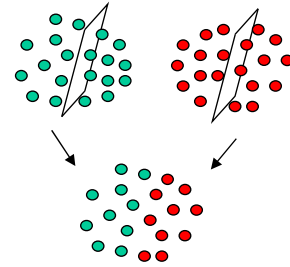D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, and T. R. Walsh, Adv. Chem. Phys. **115**, (2000)

Z. Li, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **84**, 6611 (1987)

further methods...

- Genetic Algorithms[1]:

    ⟹   two "parents" mate and create a "child"

    ⟹   the fittest (energetically lowest) children "survive"

⟹ | talk by Matt Probert, Friday, 26th June, 9:00, "Structure predictions in materials science"

- Minima-Hopping, Basin-Paving, Conformational Space Annealing,...

[1]D. M. Deaven, K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995)

Summary

- How to get E_{tot} in DFT
    - density-mixing schemes (Pulay, Broyden)
    - direct minimization

- Taking the derivative of E_{tot}
    - atomic forces
    - molecular vibrations

- What to do with E_{tot} and its derivative
    - local structural relaxation (steepest-descent, conjugate gradient, BFGS)
    - global structure optimization (Basin-Hopping, Genetic Algorithms,...)

Summary

- How to get E_{tot} in DFT
  - density-mixing schemes (Pulay, Broyden)
  - direct minimization

- Taking the derivative of E_{tot}
  - atomic forces
  - molecular vibrations

- What to do with E_{tot} and its derivative
  - local structural relaxation (steepest-descent, conjugate gradient, BFGS)
  - global structure optimization (Basin-Hopping, Genetic Algorithms,...)

Let's do it !!

today, 14:00 – 18:00
first practical session
lecture hall