University at Buffalo
The State University of New York

# *Excited States and Nonadiabatic Dynamics*
# *CyberTraining School/Workshop 2022*

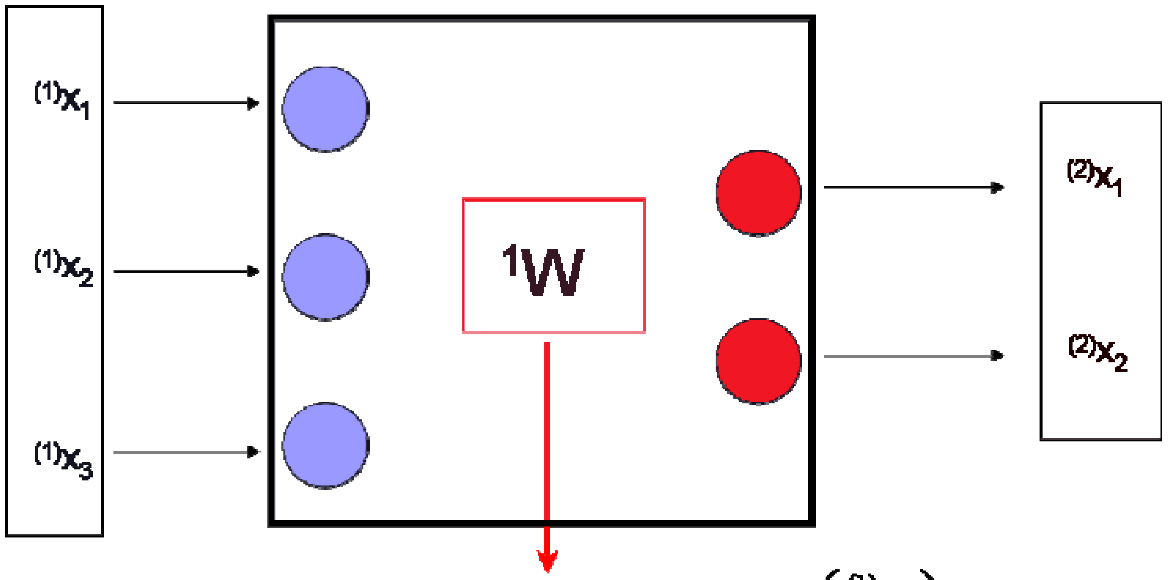
**Alexey Akimov**

University at Buffalo, SUNY

July 4, 2022

# *Fundamentals of ML with Libra*

# Basics of the ANN




$$\begin{pmatrix} {}^{(2)}NET_1 \\ {}^{(2)}NET_2 \end{pmatrix} = \begin{pmatrix} {}^{(1)}w_{11} & {}^{(1)}w_{12} & {}^{(1)}w_{13} \\ {}^{(1)}w_{21} & {}^{(1)}w_{22} & {}^{(1)}w_{23} \end{pmatrix} \begin{pmatrix} {}^{(1)}x_1 \\ {}^{(1)}x_2 \\ {}^{(1)}x_3 \end{pmatrix}$$

Transfer function:

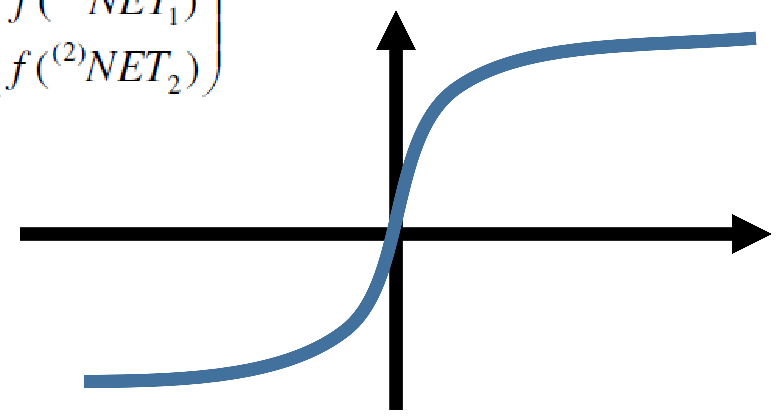
$$f(x) = \tanh(\alpha x)$$

$$f(x) = \frac{1}{1 + \exp(-\alpha x)}$$

- The expected output should be scaled
- The inputs should be close to 0.0, where the transfer function slope is large

$$\begin{pmatrix} {}^{(2)}NET_1 \\ {}^{(2)}NET_2 \end{pmatrix} = \begin{pmatrix} {}^{(1)}w_{11} & {}^{(1)}w_{12} & {}^{(1)}w_{13} \\ {}^{(1)}w_{21} & {}^{(1)}w_{22} & {}^{(1)}w_{23} \end{pmatrix} \begin{pmatrix} {}^{(1)}x_1 \\ {}^{(1)}x_2 \\ {}^{(1)}x_3 \end{pmatrix} + \begin{pmatrix} {}^{(1)}b_1 \\ {}^{(1)}b_2 \\ {}^{(1)}b_3 \end{pmatrix}$$

$$\begin{pmatrix} {}^{(2)}x_1 \\ {}^{(2)}x_2 \end{pmatrix} = \begin{pmatrix} f({}^{(2)}NET_1) \\ f({}^{(2)}NET_2) \end{pmatrix}$$



# Function back-propagation

University at Buffalo
The State University of New York

$$y_i^{(L)} = f\left(\sum_j W_{ij}^{(L)} y_j^{(L-1)} + B_i^{(L)}\right) \leftrightarrow \boldsymbol{y}^{(L)} = \boldsymbol{f}(\boldsymbol{W}^{(L)}\boldsymbol{y}^{(L-1)} + \boldsymbol{B}^{(L)})$$

$$\boldsymbol{a}^{(L)} = \boldsymbol{W}^{(L)}\boldsymbol{y}^{(L-1)} + \boldsymbol{B}^{(L)}$$

$$\frac{\partial y_i^{(L)}}{\partial W_{ab}^{(L)}} = f'\left(a_i^{(L)}\right)\sum_j \left(\frac{\partial W_{ij}^{(L)}}{\partial W_{ab}^{(L)}} y_j^{(L-1)}\right) = f'\left(a_i^{(L)}\right)\sum_j \left(\delta_{ia}\delta_{jb} y_j^{(L-1)}\right) = \delta_{ia} f'\left(a_i^{(L)}\right) y_b^{(L-1)}$$

$$\frac{\partial y_i^{(L)}}{\partial B_a^{(L)}} = f'\left(a_i^{(L)}\right)\frac{\partial B_i^{(L)}}{\partial B_a^{(L)}} = \delta_{ia} f'\left(a_i^{(L)}\right)$$

$$E = \frac{1}{2}\sum\nolimits_i \left(y_i^{(NL)} - t_i\right)^2 \leftrightarrow E = \frac{1}{2}\left(\boldsymbol{y}^{(NL)} - \boldsymbol{t}\right)^T\left(\boldsymbol{y}^{(NL)} - \boldsymbol{t}\right)$$

## Gradients w.r.t the weights and biases of the layer NL

University at Buffalo
The State University of New York

$$\frac{\partial E}{\partial W_{ab}^{(NL)}} = \sum_i \left(y_i^{(NL)} - t_i\right) f'\left(a_i^{(NL)}\right) \sum_j y_j^{(NL-1)} \delta_{ai}\delta_{bj} = \left(y_a^{(NL)} - t_a\right) f'\left(a_a^{(NL)}\right) y_b^{(NL-1)}$$

$$\frac{\partial E}{\partial B_a^{(NL)}} = \sum_i \left(y_i^{(NL)} - t_i\right) f'\left(a_i^{(NL)}\right) \delta_{ia} = \left(y_a^{(NL)} - t_a\right) f'\left(a_a^{(NL)}\right)$$

$$d_a^{(NL)} = \left(y_a^{(NL)} - t_a\right) f'\left(a_a^{(NL)}\right)$$

Introducing matrix $\boldsymbol{\tau}^{(NL)}$ - the diagonal matrix, with $\tau_{ii}^{(NL)} = f'\left(a_i^{(NL)}\right)$

$$\boldsymbol{d}^{(NL)} = \boldsymbol{\tau}^{(NL)}\left(\boldsymbol{y}^{(NL)} - \boldsymbol{t}\right)$$

$$\frac{\partial E}{\partial W_{ab}^{(NL)}} = d_a^{(NL)} y_b^{(NL-1)} \leftrightarrow \frac{\partial E}{\partial \boldsymbol{W}^{(\boldsymbol{NL})}} = \left(\boldsymbol{d}^{(NL)}\right)^T \boldsymbol{y}^{(NL-1)}$$

$$\frac{\partial E}{\partial B_a^{(NL)}} = d_a^{(NL)} \leftrightarrow \frac{\partial E}{\partial \boldsymbol{B}^{(NL)}} = \boldsymbol{d}^{(NL)}$$

# In general

University at Buffalo
The State University of New York

$$\frac{\partial E}{\partial W_{ab}^{(NL-1)}} = d_a^{(NL-1)} y_b^{(NL-2)} \leftrightarrow \frac{\partial E}{\partial \boldsymbol{W}^{(\boldsymbol{NL-1})}} = \left(\boldsymbol{d}^{(NL-1)}\right)^T \boldsymbol{y}^{(NL-2)}$$

$$\frac{\partial E}{\partial B_a^{(NL-1)}} = d_a^{(NL-1)} \leftrightarrow \frac{\partial E}{\partial \boldsymbol{B}^{(NL-1)}} = \boldsymbol{d}^{(NL-1)}$$

# Derivatives of the ANN w.r.t. Inputs

University at Buffalo
The State University of New York

$$y_i^{(L)} = f\left(\sum_j W_{ij}^{(L)} y_j^{(L-1)} + B_i^{(L)}\right) \leftrightarrow \boldsymbol{y}^{(L)} = \boldsymbol{f}(\boldsymbol{W}^{(L)}\boldsymbol{y}^{(L-1)} + \boldsymbol{B}^{(L)})$$

$$\boldsymbol{a}^{(L)} = \boldsymbol{W}^{(L)}\boldsymbol{y}^{(L-1)} + \boldsymbol{B}^{(L)}$$

$$\frac{\partial y_i^{(L)}}{\partial y_a^{(0)}} = f'\left(a_i^{(L)}\right) \sum_j \left( W_{ij}^{(L)} \frac{\partial y_j^{(L-1)}}{\partial y_a^{(0)}} \right)$$

$$= f'\left(a_i^{(L)}\right) \sum_j \left( W_{ij}^{(L)} f'\left(a_j^{(L-1)}\right) \sum_k W_{jk}^{(L-1)} \frac{\partial y_k^{(L-2)}}{\partial y_a^{(0)}} \right)$$

$$= \sum_{j,k} f'\left(a_i^{(L)}\right) W_{ij}^{(L)} f'\left(a_j^{(L-1)}\right) W_{jk}^{(L-1)} \frac{\partial y_k^{(L-2)}}{\partial y_a^{(0)}} = \cdots$$

$$\frac{\partial y_i^{(L)}}{\partial y_a^{(0)}} = G_{ia}^{(L)}$$

$$\boldsymbol{G}^{(L)} = \boldsymbol{\tau}^{(L)} \boldsymbol{W}^{(L)} \boldsymbol{G}^{(L-1)}$$

More specifically,

$$\boldsymbol{G}^{(NL)} = \boldsymbol{\tau}^{(NL)} \boldsymbol{W}^{(NL)} \boldsymbol{\tau}^{(NL-1)} \boldsymbol{W}^{(NL-1)} \boldsymbol{\tau}^{(NL-2)} \boldsymbol{W}^{(NL-2)} \ldots \boldsymbol{\tau}^{(1)} \boldsymbol{W}^{(1)}$$

# Back-prop with momentum

http://page.mi.fu-berlin.de/rojas/neural/chapter/K8.pdf

https://arxiv.org/pdf/1711.05101.pdf

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.1332

$$\Delta w_k(i) = -\gamma \frac{\partial E}{\partial w_k} + \alpha \Delta w_k(i-1),$$

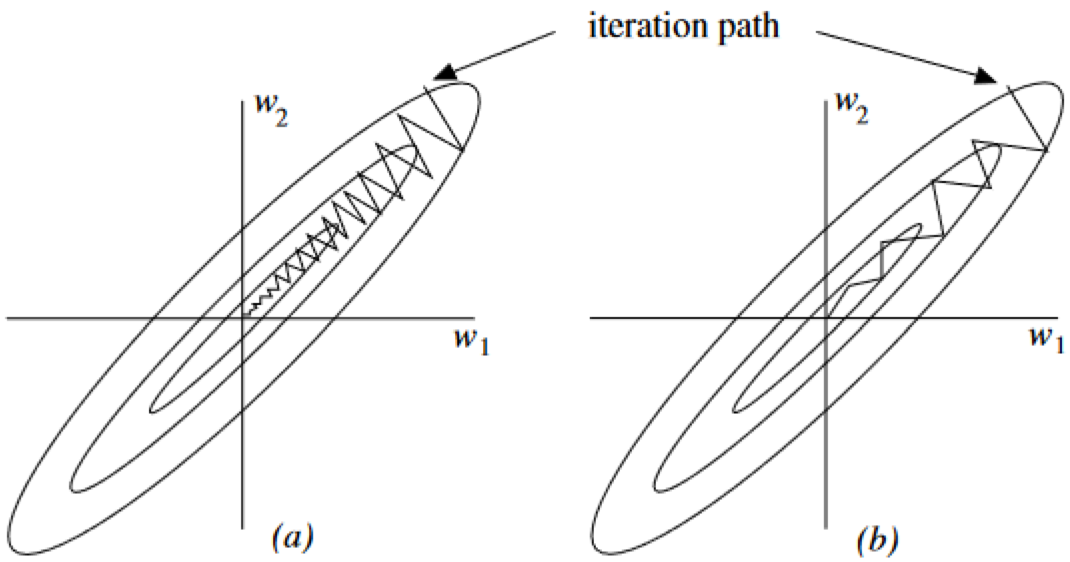



**Fig. 8.1.** Backpropagation without (a) or with (b) momentum term

# Rprop

University at Buffalo
The State University of New York

$$\gamma_i^{(k+1)} = \begin{cases} \min(\gamma_i^{(k)} u, \gamma_{max}) & \text{if } \nabla_i E^{(k)} \cdot \nabla_i E^{(k-1)} > 0 \\ \max(\gamma_i^{(k)} d, \gamma_{min}) & \text{if } \nabla_i E^{(k)} \cdot \nabla_i E^{(k-1)} < 0 \\ \gamma_i^{(k)} & \text{otherwise,} \end{cases}$$

where the constants $u$ and $d$ satisfy $u > 1$ and $d < 1$, as usual. When $\nabla_i E^{(k)} \cdot \nabla_i E^{(k-1)} \geq 0$ the weight updates are given by

$$\Delta^{(k)} w_i = -\gamma_i^{(k)} \mathrm{sgn}(\nabla_i E^{(k)}),$$

following rule:

$$\Delta_{ij}^{(t)} := \begin{cases} \eta^+ \cdot \Delta_{ij}^{(t-1)} & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)} \cdot \frac{\partial E}{\partial w_{ij}}^{(t)} > 0 \\ \eta^- \cdot \Delta_{ij}^{(t-1)} & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)} \cdot \frac{\partial E}{\partial w_{ij}}^{(t)} < 0 \\ \Delta_{ij}^{(t-1)} & , \text{ else }, \end{cases} \tag{1}$$

where $0 < \eta^- < 1 < \eta^+$. If the partial deriva-

# *Fundamentals of DVR with Libra*

# Wavefunction is discretized on a grid

University at Buffalo
The State University of New York

Wavefunction is discretized on a grid

$$\langle r|\Psi\rangle = \Psi(r,t) = \sum_{\substack{i\in grid,\\ a}} \Psi_a(r_i,t)\delta(r-r_i)|a\rangle$$

$$\text{PSI\_dia} = \underbrace{\left\{\begin{pmatrix}\Psi_0(r_0)\\ \dots \\ \Psi_{N-1}(r_0)\end{pmatrix}, \begin{pmatrix}\Psi_0(r_1)\\ \dots \\ \Psi_{N-1}(r_1)\end{pmatrix}, \dots, \begin{pmatrix}\Psi_0(r_{Npts-1})\\ \dots \\ \Psi_{N-1}(r_{Npts-1})\end{pmatrix}\right\}}$$

In Libra, any N-dimensional grid is "linearized" this way via a mapping function

This could be thought of as using the basis of grid-point functions $|i,a\rangle$: $\langle r|\Psi\rangle = \delta(r-r_i)|a\rangle$

Overlaps

$$\langle\Psi|\Psi\rangle = \sum_{a,b,i,j} \int dr \Psi_a^*(r_i)\Psi_b(r_j)\delta(r-r_i)\delta(r-r_j)\langle a|b\rangle = \Delta r \sum_{a,i} \Psi_a^*(r_i)\Psi_a(r_i)$$

Matrix elements of operators

$$\langle\Psi|\hat{A}|\Psi\rangle = \sum_{a,b,i,j} \int dr \Psi_a^*(r_i)\Psi_b(r_j)\delta(r-r_i)\delta(r-r_j)A_{ab}(r) = \Delta r \sum_{a,b,i} \Psi_a^*(r_i)A_{ab}(r_i)\Psi_a(r_i)$$

# Momentum representation

Real-space (coordinate) wavefunction

$$\psi_a(\boldsymbol{r},t)=\int\tilde{\psi}_a(\boldsymbol{k},t)e^{2\pi i\boldsymbol{r}\boldsymbol{k}}d\boldsymbol{k}$$

Reciprocal-space (momentum) wavefunction

$$\tilde{\psi}_i(\boldsymbol{k},t)=\int\psi_i(\boldsymbol{r},t)e^{-2\pi i\boldsymbol{r}\boldsymbol{k}}d\boldsymbol{r}$$

$$\left\langle\psi_i(x)\middle|\left(-i\frac{\partial}{\partial x}\right)^n\middle|\psi_j(x)\right\rangle=\sum_{i,j}\int dx\left(\int\tilde{\psi}_i(k)e^{2\pi ixk}dk\right)\left(-i\frac{\partial}{\partial x}\right)^n\left(\int\tilde{\psi}_j(k')e^{2\pi ixk'}dk'\right)$$

$$=(-i)^n\sum_{i,j}\int dx\left(\int\tilde{\psi}_i(k)e^{2\pi ixk}dk\right)^*\left((2\pi i)^n\int k'^n\tilde{\psi}_j(k')e^{2\pi ixk'}dk'\right)$$

$$=(2\pi)^n\sum_{i,j}\int dxdkdk'\tilde{\psi}_i^*(k)e^{-2\pi ixk}(k')^n\tilde{\psi}_j(k')e^{2\pi ixk'}$$

$$=(2\pi)^n\sum_{i,j}\int dkdk'\tilde{\psi}_i^*(k)\delta(k-k')(k')^n\tilde{\psi}_j(k')=(2\pi)^n\sum_{i,j}\int dk\tilde{\psi}_i^*(k)k^n\tilde{\psi}_j(k)$$

$$\to(2\pi)^n\Delta k\sum_{i,j,m}\tilde{\psi}_i^*(k_m)k_m^n\tilde{\psi}_j(k_m)$$

# Solution of the TD-SE

$$i\hbar \frac{\partial \Psi(r,t)}{\partial t} = \hat{H}\Psi(r,t) = (\hat{T} + \hat{V})\Psi(r,t)$$

Finite difference evaluation of the derivatives

$$\partial_t \Psi_i(\boldsymbol{r}_n, t_m) = \frac{1}{2\Delta t}[\Psi_i(\boldsymbol{r}_n, t_{m+1}) - \Psi_i(\boldsymbol{r}_n, t_{m-1})]$$

$$\nabla_{\boldsymbol{r}_\alpha} \Psi_i(x_n, t_m) = \frac{1}{2\Delta r_\alpha}\left[\Psi_i(\boldsymbol{r}_{\alpha,n+1}, t_m) - \Psi_i(\boldsymbol{r}_{\alpha,n-1}, t_m)\right]$$

$$\nabla^2_{\boldsymbol{r}_\alpha} \Psi_i(x_n, t_m) = \frac{1}{4\Delta \boldsymbol{r}_\alpha^2}\left[\Psi_i(\boldsymbol{r}_{\alpha,n+2}, t_m) - \Psi_i(\boldsymbol{r}_n, t_m) - [\Psi_i(\boldsymbol{r}_n, t_m) - \Psi_i(\boldsymbol{r}_{\alpha,n-2}, t_m)]\right] = \frac{1}{4\Delta r_\alpha^2}\left[\Psi_i(\boldsymbol{r}_{\alpha,n+2}, t_m) - 2\Psi_i(\boldsymbol{r}_n, t_m) + \Psi_i(\boldsymbol{r}_{\alpha,n-2}, t_m)\right]$$

# Solution of the TD-SE

University at Buffalo
The State University of New York

$$i\hbar\frac{\partial|\Psi(t)\rangle}{\partial t} = \hat{H}|\Psi(t)\rangle = \left(\hat{T}+\hat{V}\right)|\Psi(t)\rangle$$

Split-operator method (Kosloff & Kosloff)

$$|\Psi(t+\Delta t)\rangle = \exp\left(-\frac{i\Delta t}{\hbar}\hat{H}\right)|\Psi(t)\rangle = \exp\left(-\frac{i\Delta t}{\hbar}\left(\hat{T}+\hat{V}\right)\right)|\Psi(t)\rangle \approx \exp\left(-\frac{i\Delta t}{2\hbar}\hat{V}\right)\exp\left(-\frac{i\Delta t}{\hbar}\hat{T}\right)\exp\left(-\frac{i\Delta t}{2\hbar}\hat{V}\right)|\Psi(t)\rangle$$

$$\Psi_a(r_i,t') = \langle r_i,a|\exp\left(-\frac{i\Delta t}{2\hbar}\hat{V}\right)|\Psi(t)\rangle = \langle r_i|\exp\left(-\frac{i\Delta t}{2\hbar}\hat{V}\right)\sum_{j,b}|r_j,b\rangle\langle r_j,b|\Psi(t)\rangle = \sum_{j,b}\left\langle r_i,a\right|\exp\left(-\frac{i\Delta t}{2\hbar}\hat{V}\right)\left|r_j,b\right\rangle\langle r_j,b|\Psi(t)\rangle$$

$$= \sum_{j,b}\left\langle a\right|\exp\left(-\frac{i\Delta t}{2\hbar}V\left(r_i\right)\right)\left|b\right\rangle\delta_{ij}\Psi_b\left(r_j,t\right) = \sum_{b}\left[\exp\left(-\frac{i\Delta t}{2\hbar}V\left(r_i\right)\right)\right]_{ab}\Psi_b(r_i,t)$$

$$\tilde{\Psi}_{\mathrm{a}}(k_i,t'') = \langle k_i,a|\exp\left(-\frac{i\Delta t}{\hbar}\hat{T}\right)|\Psi(t)\rangle = \langle k_i,a|\exp\left(-\frac{i\Delta t}{\hbar}\hat{T}\right)\sum_{j,b}|k_j,b\rangle\langle k_j,b|\Psi(t)\rangle$$

$$= \sum_{j,b}\left\langle k_i,a\right|\exp\left(-\frac{i\Delta t}{2\hbar}\hat{T}\right)\left|k_j,b\right\rangle\left\langle k_j,b\right|\Psi(t)\rangle = \sum_{j,b}\exp\left(-\frac{i\Delta t}{2\hbar}\frac{k_i^2}{2m}\right)\delta_{ij}\delta_{ab}\tilde{\Psi}_b(t) = \exp\left(-\frac{i\Delta t}{2\hbar}\frac{k_i^2}{2m}\right)\tilde{\Psi}_a(t)$$

# *Fundamentals of HEOM with Libra*

See here

https://compchem-cybertraining.github.io/Cyber_Training_Workshop_2021/files/Jain-HEOM.pdf