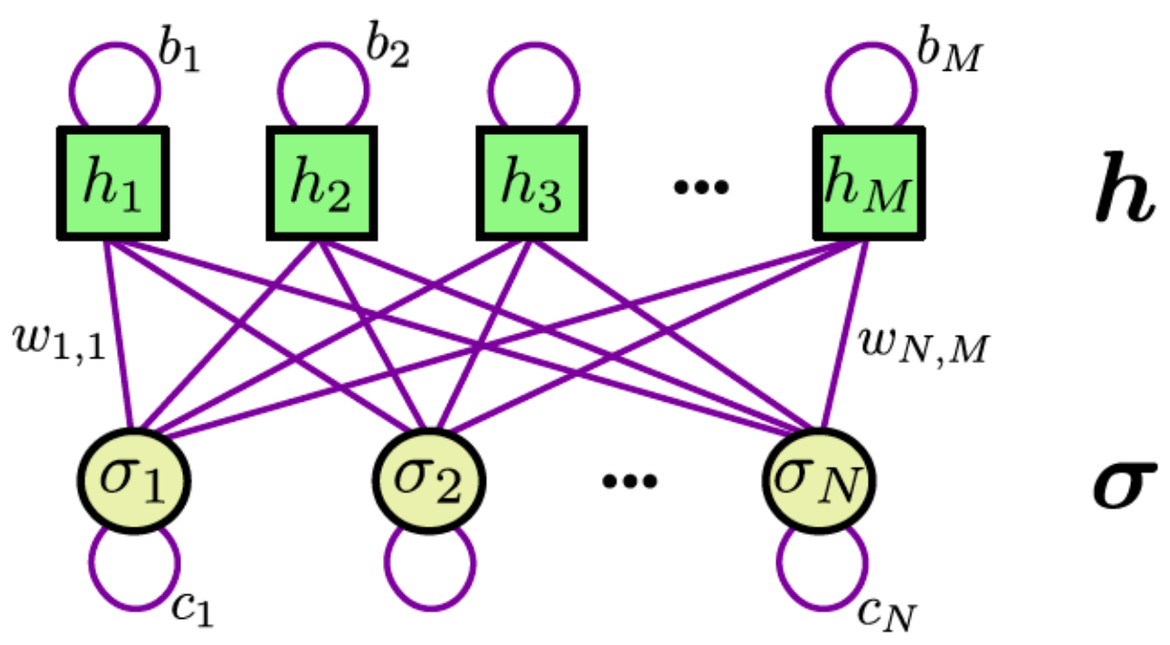



*Figure 1: Representation of a RBM network. The yellow circles denote the physical sites, the green boxes the hidden units and the purple lines the (complex) network weights that play the role of variational parameters in the VMC calculation.*

$$\mathcal{E}_{\mathrm{nw}}(\boldsymbol{\sigma}, \boldsymbol{h}; \boldsymbol{w}) = -\sum_{i=1}^{N} c_i \sigma_i + \sum_{j=1}^{M} \left( b_j + \sum_{i=1}^{N} \sigma_i w_{i,j} \right) h_j$$

$$\psi_{\boldsymbol{w}}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{h}} \mathrm{e}^{-\mathcal{E}_{\mathrm{nw}}(\boldsymbol{\sigma}, \boldsymbol{h}; \boldsymbol{w})} = \mathrm{e}^{\sum_{i=1}^{N} c_i \sigma_i} \prod_{j=1}^{M} 2 \cosh \theta_{\boldsymbol{w}}^{(j)}(\boldsymbol{\sigma})$$

$$\theta_{\boldsymbol{w}}^{(j)}(\boldsymbol{\sigma}) \equiv b_j + \sum_{i=1}^{N} \sigma_i w_{i,j}$$

# Solutions – Sheet 6

## 1.) Local energy estimator

$$E_{loc}(\sigma) = \frac{\langle \sigma | H | \Psi_w \rangle}{\langle \sigma | \Psi_w \rangle} \quad ; \quad \boxed{\sigma_j = \pm 1}$$

- Problem: Neither $|\sigma\rangle$ nor $|\Psi_w\rangle$ is an eigenstate of $H$.
- Solution: Compute the action of $H$ on $|\sigma\rangle$:

$$H|\sigma\rangle = \left( -\frac{J}{2} \sum_j \left( S_j^+ S_{j+1}^- + \text{h.c.} \right) + \sum_j \Delta S_j^z S_{j+1}^z \right) |\sigma\rangle =$$

$$= \left( \Delta \sum_j \frac{\sigma_j}{2} \frac{\sigma_{j+1}}{2} \right) |\sigma\rangle - \frac{J}{2} \sum_j \frac{(1-\sigma_j)}{2} \frac{(1+\sigma_{j+1})}{2} \Big| \sigma_1 \dots \overbrace{\sigma_j}^{\text{flipped}}, \overbrace{\sigma_{j+1}}^{\text{flipped}} \dots \sigma_N \Big\rangle$$

$$-\frac{J}{2}\sum_j \frac{(1+\sigma_j)}{2}\frac{(1-\sigma_{j+1})}{2}\,|\sigma_1,\dots,\bar{\sigma}_j,\bar{\sigma}_{j+1},\dots\sigma_N\rangle,$$

i.e. $H|\sigma\rangle = \sum_{\sigma'} h_{\sigma'\sigma}|\sigma'\rangle$, where the sum on $\sigma'$ only runs on $O(N)$ terms, not over the entire Hilbert space.

$$\rightsquigarrow \quad \boxed{E_{loc}(\sigma) = \frac{\sum_{\sigma'} h_{\sigma'\sigma}\,\Psi_w(\sigma')}{\Psi_w(\sigma)}}$$

## 2. Gradients in parameter space

- $O^{(\sigma)}_{c_i} := \frac{\partial \log \Psi_w(\sigma)}{\partial(\mathrm{Re}\,c_i)} = \frac{1}{\Psi_w(\sigma)}\frac{\partial \Psi_w(\sigma)}{\partial(\mathrm{Re}\,c_i)} = \sigma_i$

($c_i$: real part)

- $O_{b_j^r}(\sigma) := \dfrac{\partial \log \Psi_w(\sigma)}{\partial b_j^r} = \tanh\left(\theta_w^{(i)}(\sigma)\right)$

- $O_{w_{ij}^r} := \dfrac{\partial \log \Psi_w(\sigma)}{\partial w_{ij}^r} = \sigma_i \tanh\left(\theta_w^{(i)}(\sigma)\right)$

↳ Imaginary parts. For all parameters $z_k = z_k^r + i z_k^i$, we simply get $O_{z_k}^i = i O_{z_k}^r$, $z = c, b, w$.

- Motivation for using logarithmic derivatives:

$$\frac{\langle \partial_{z_k^{r,i}} \Psi_w | \Psi_w \rangle}{\langle \Psi_w | \Psi_w \rangle} = \frac{1}{\langle \Psi_w | \Psi_w \rangle} \sum_\sigma \langle \partial_{z_k^{r,i}} \Psi_w | \sigma \rangle \langle \sigma | \Psi_w \rangle =$$

$P_w(\sigma) = \dfrac{|\Psi_w(\sigma)|^2}{\langle \Psi_w | \Psi_w \rangle}$

$$= \sum_\sigma P_w(\sigma) \frac{\langle \partial_{z_k^{r,i}} \Psi_w | \sigma \rangle}{\langle \Psi_w | \sigma \rangle} = \sum_\sigma P_w(\sigma) \, O_{z_k^{r,i}}^*(\sigma)$$

↝ Logarithmic derivatives naturally occur in gradients with respect to the variational parameters.

↳ Most important for our purposes:

$$\frac{\partial E_\omega}{\partial z_k^\mu} = \frac{\langle \partial_{z_k^\mu} \Psi_\omega | H | \Psi_\omega \rangle}{\langle \Psi_\omega | \Psi_\omega \rangle} - \frac{\langle \partial_{z_k^\mu} \Psi_\omega | \Psi_\omega \rangle}{\langle \Psi_\omega | \Psi_\omega \rangle} \cdot$$

$$\cdot \frac{\langle \Psi_\omega | H | \Psi_\omega \rangle}{\langle \Psi_\omega | \Psi_\omega \rangle} + \text{c.c.}$$

$$=: 2\,\mathrm{Re}\left( F^\omega_{z_k^\mu} \right) ,$$

with $F^{\omega}_{z_k^\mu} = \frac{1}{\langle \Psi_\omega | \Psi_\omega \rangle} \left( \langle \partial_{z_k^\mu} \Psi_\omega | H | \Psi_\omega \rangle - \langle \partial_{z_k^\mu} \Psi_\omega | \Psi_\omega \rangle E_\omega \right)$.

Using the logarithmic derivatives, we get:

$$\frac{\langle \partial_{z_k^\mu} \Psi_\omega | H | \Psi_\omega \rangle}{\langle \Psi_\omega | \Psi_\omega \rangle} = \sum_\sigma \underbrace{\frac{|\Psi_\omega(\sigma)|^2}{\langle \Psi_\omega | \Psi_\omega \rangle}}_{P_\omega(\sigma)} \underbrace{\frac{\langle \partial_{z_k^\mu} \Psi_\omega | \sigma \rangle}{\langle \Psi_\omega | \sigma \rangle}}_{O^*_{z_k^\mu}} \underbrace{\frac{\langle \sigma | H | \Psi_\omega \rangle}{\langle \sigma | \Psi_\omega \rangle}}_{= E_{loc}(\sigma)}$$

$$= \sum_\sigma P_\omega(\sigma)\, O^*_{z_k^\mu}\, E_{loc}(\sigma), \text{ and}$$

$$\text{overall } F^{\omega}_{z_k^\mu} = \sum_\sigma O^*_{z_k^\mu}(\sigma) E_{loc}(\sigma) P_\omega(\sigma) - \left( \sum_\sigma P_\omega(\sigma)\, O^*_{z_k^\mu}(\sigma) \right) \left( \sum_{\sigma'} P_\omega(\sigma') E_{loc}(\sigma') \right)$$

$$= \langle O^*_{z_k^\mu} E_{loc} \rangle - \langle O^*_{z_k^\mu} \rangle \langle E_{loc} \rangle,$$

where $\langle \dots \rangle$ means ensemble average.

$\rightsquigarrow$ This expression is ready for Metropolis sampling, replacing all $\sum_{\sigma} P_\omega(\sigma) X(\sigma) \longrightarrow \frac{1}{M_s} \sum_{e=1}^{M_s} X(\sigma_e)$.

## 3. Metropolis sampling

Simplest proposal matrix: Pick random pair of spins and propose a flip-flop, i.e. $|\sigma'\rangle = |\sigma_1, \dots, \bar{\sigma}_j, \dots \bar{\sigma}_e, \dots\rangle$ with $\sigma_j \neq \sigma_e$.

$\rightsquigarrow$ $\mathrm{Prop}(\sigma'|\sigma) = \begin{cases} \text{const}, & \text{if } \sigma' \text{ and } \sigma \text{ differ by single flip-flop} \\ 0 & \text{otherwise.} \end{cases}$

$\rightsquigarrow$ this preserves magnetization (as does the Hamiltonian), and is symmetric, i.e $\mathrm{Prop}(\sigma'|\sigma) = \mathrm{Prop}(\sigma|\sigma')$

↳ $P_{acc}(\sigma'|\sigma) = \min\left\{1, \frac{|\Psi(\sigma')|^2}{|\Psi(\sigma)|^2}\right\}$.

↑ Metropolis

↳ Next level of sophistication: $P_{prop} \sim e^{-\frac{|\vec{v}-\vec{e}|}{\xi}}$ (finite corr. length)

( reflects locality of the Hamiltonian )

- Generate sample for walkers starting from random initial $\sigma_i$ (with desired magnetization), and add $\sigma_{\ell \cdot L_e}$ to ensemble every $L_e$ steps, where $L_e$ (here $\mathcal{O}(N)$ is the equilibration time, i.e. ensemble consists of $\{\sigma_{L_e}, \sigma_{2L_e}, \dots \sigma_{M_s L_e}\}$.

→ See Toulouse et al. arXiv: 1508.02989

↳ Several walkers may contribute in parallel to the ensemble.

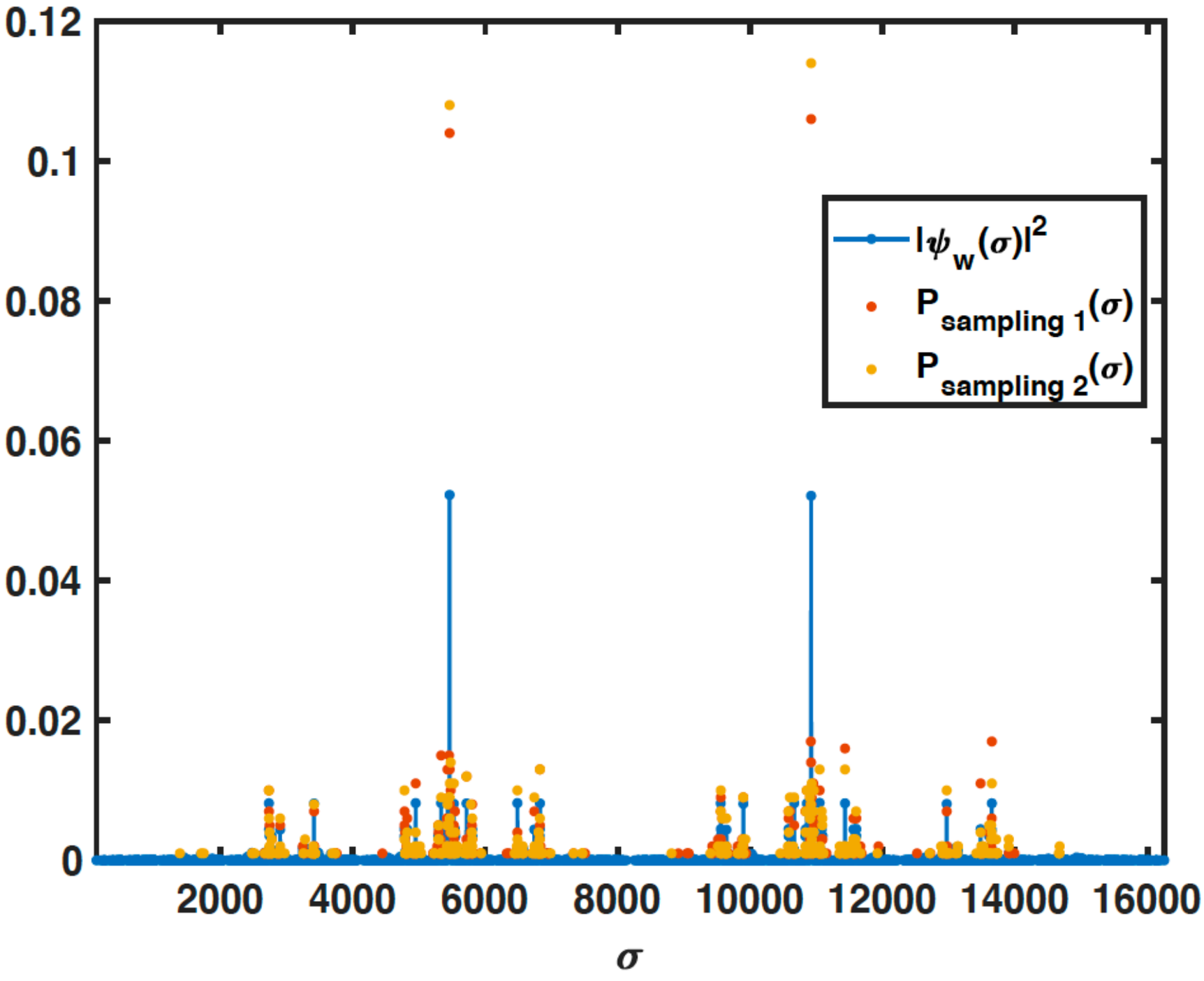



*Figure 2: Example of a comparison between* $P_{\boldsymbol{w}}(\boldsymbol{\sigma}) \equiv \frac{|\langle \boldsymbol{\sigma}|\psi_{\boldsymbol{w}}\rangle|^2}{\langle \psi_{\boldsymbol{w}}|\psi_{\boldsymbol{w}}\rangle}$ *(blue data points) calculated for a RBM state approximating the ground-state for a lattice of* $N = 14$ *sites with PBC,* $J = 1$ *and* $\Delta = 1$*, and the sampling probabilities from two instances of sampling (orange and yellow data points).*

4. Stochastic gradient optimization of variational parameters

- Update iteratively $\delta w = -\gamma \underbrace{\nabla_w \bar{E}_w}_{\sim \mathrm{Re}(\vec{F})}$, (step size: $\gamma$)

where $\nabla_w E_w$ is estimated using Metropolis sampling in terms of the force vector (see subproblem 2).

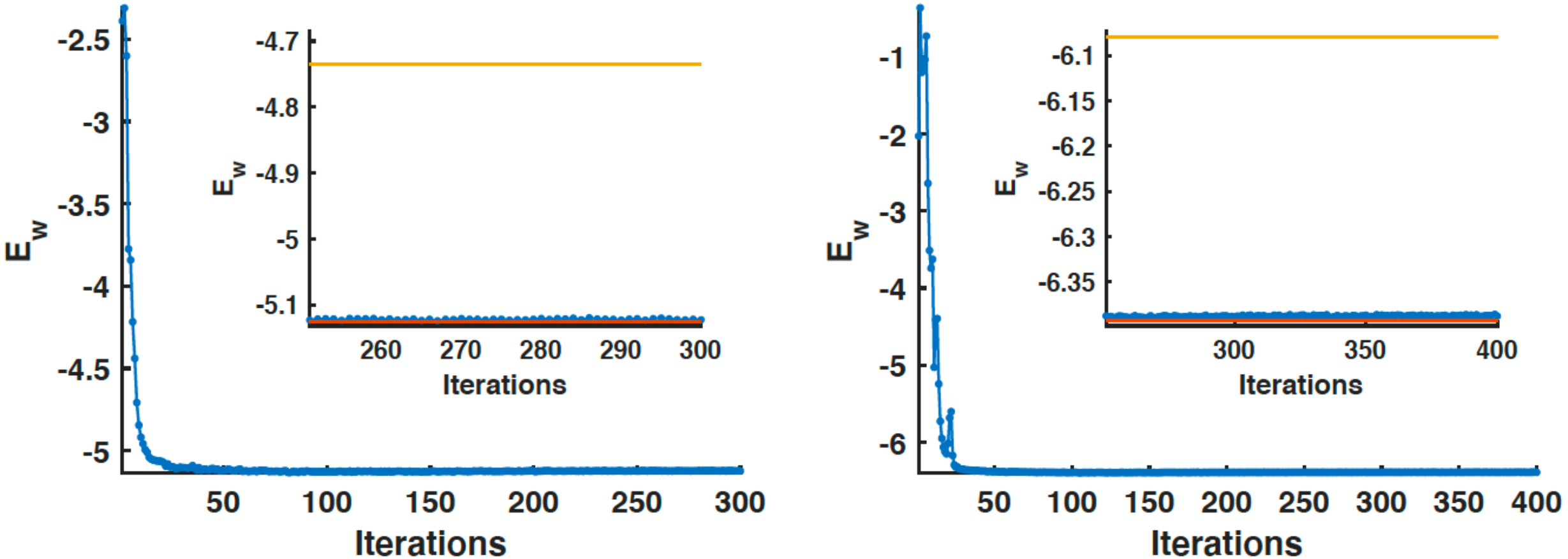



*Figure 3: Example of a convergence plots of $E_{\boldsymbol{w}}$ (blue data points) during the optimization iterations. The first data point in the main plots corresponds to the energy calculated from a RBM state with randomly initialized weights (real parts uniformly distributed in $[-0.1, 0.1]$ and imaginary parts in $[-0.1\,\pi, 0.1\,\pi]$ ). The following data points correspond to the energies after each update . In the insets the later optimization steps are shown, compared with the exact ground-state energy (red line) and the energy of the first excited state (yellow line) obtained from exact diagonalization. Both panels refer to a lattice of with PBC, with $J = 1$ and $\Delta = 0$. On the left $N = 16$ and on the right $N = 20$. The learning rate $\gamma = 0.1$ was kept constant, and at each iteration 2000 samples for $N = 16$ and 4000 for $N = 20$ were generated.*

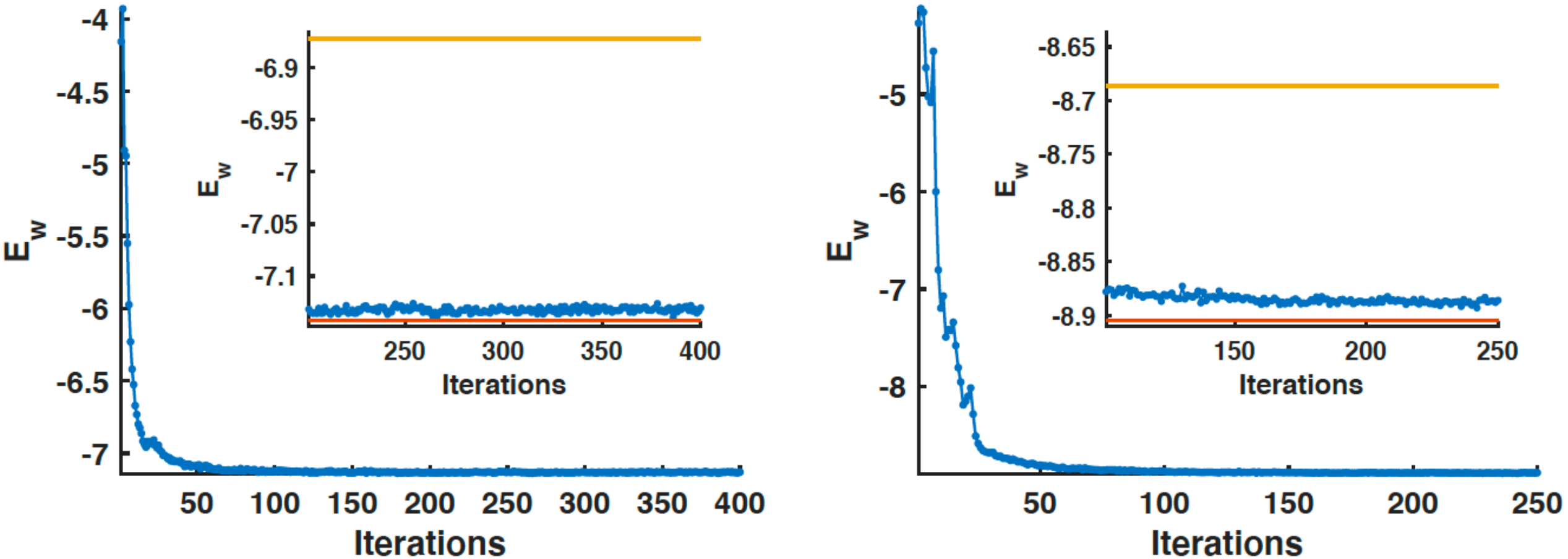
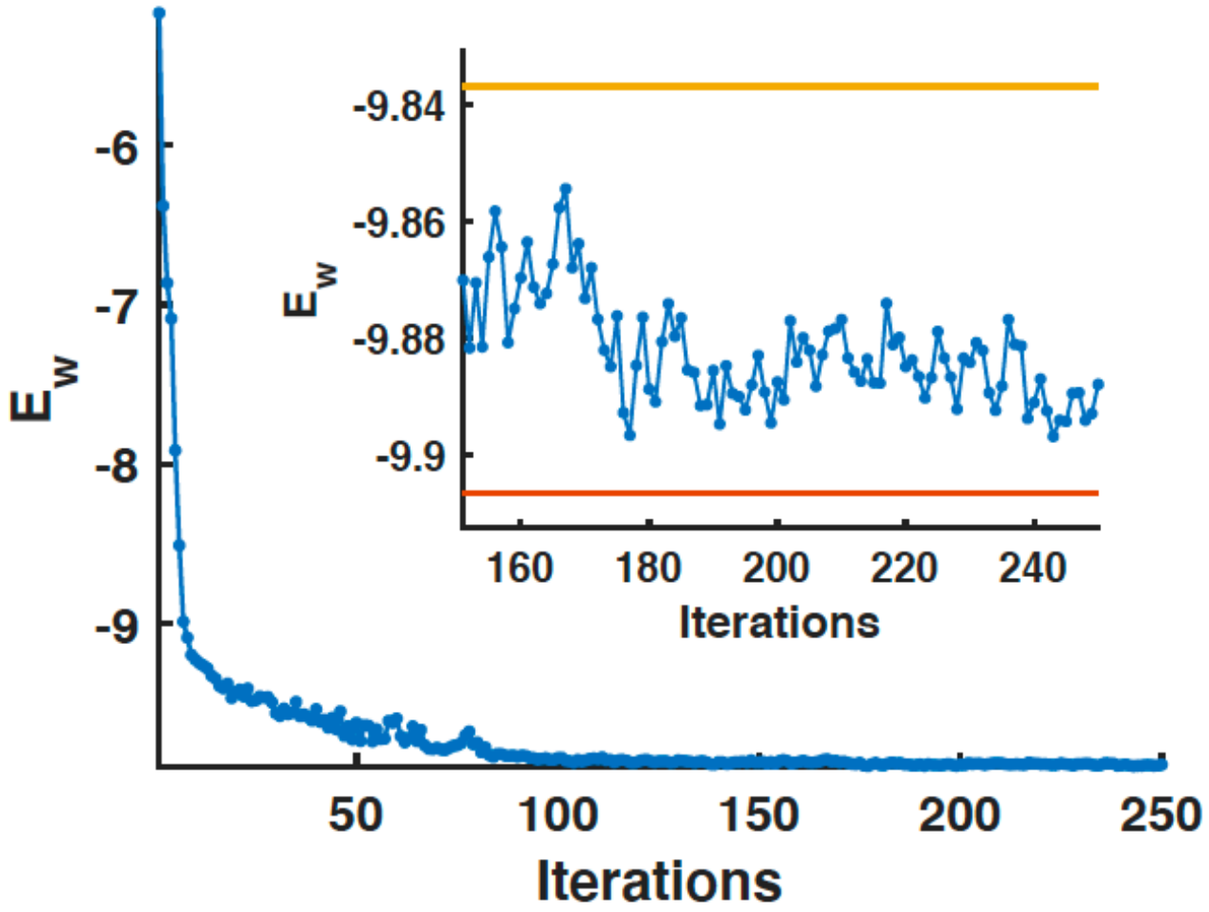



*Figure 4: Example of a convergence plots of $E_{\boldsymbol{w}}$ (blue data points) during the optimization iterations. The first data point in the main plots corresponds to the energy calculated from a RBM state with randomly initialized weights (real parts uniformly distributed in $[-0.1, 0.1]$ and imaginary parts in $[-0.1\,\pi, 0.1\,\pi]$ ). The following data points correspond to the energies after each update . In the insets the later optimization steps are shown, compared with the exact ground-state energy (red line) and the energy of the first excited state (yellow line) obtained from exact diagonalization. Both panels refer to a lattice of with PBC, with $J = 1$ and $\Delta = 1$. On the left $N = 16$ and on the right $N = 20$. The learning rate $\gamma = 0.1$ was kept constant, and at each iteration 2000 samples for $N = 16$ and 4000 for $N = 20$ were generated.*



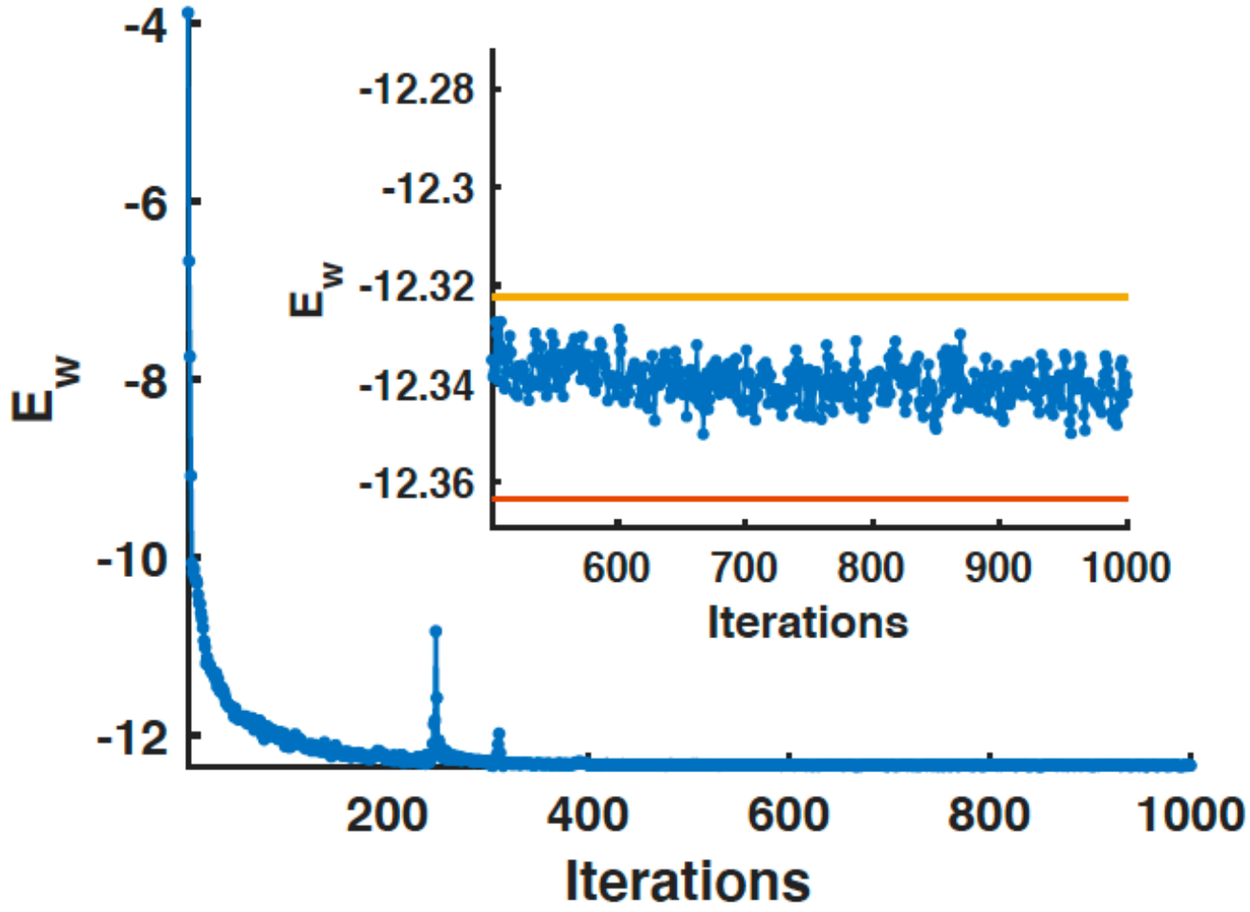




*Figure 5: Example of a convergence plots of $E_{\boldsymbol{w}}$ (blue data points) during the optimization iterations. The first data point in the main plots corresponds to the energy calculated from a RBM state with randomly initialized weights (real parts uniformly distributed in $[-0.1, 0.1]$ and imaginary parts in $[-0.1\,\pi, 0.1\,\pi]$ ). The following data points correspond to the energies after each update . In the insets the later optimization steps are shown, compared with the exact ground-state energy (red line) and the energy of the first excited state (yellow line) obtained from exact diagonalization. Both panels refer to a lattice of with PBC, with $J = 1$ and $\Delta = 2$. On the left $N = 16$ and on the right $N = 20$. The learning rate $\gamma = 0.1$ was kept constant, and at each iteration 2000 samples for $N = 16$ and 4000 for $N = 20$ were generated.*

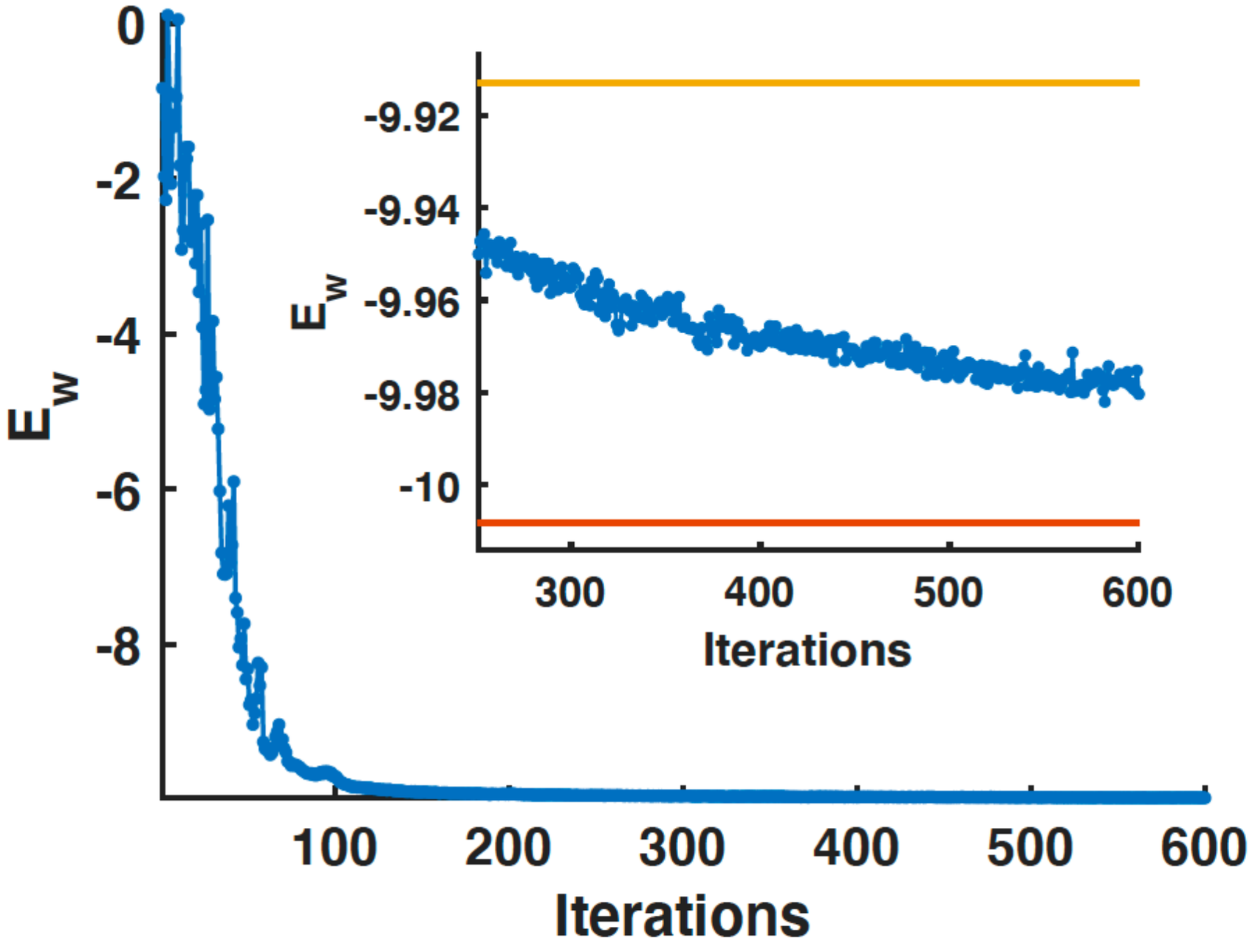



*Figure 6: Example of a convergence plot of $E_{\boldsymbol{w}}$ for $N = 32$ sites with OBC, $J = 1$ and $\Delta = 0$. In the inset the later optimization steps are shown, compared with the exact ground-state energy (red line) and the energy of the first excited state (yellow line) obtained from the Jordan-Wigner transformation of the model to free fermions in the case of $\Delta = 0$ and OBC. Here $\gamma = 0.1$ was kept constant, and at each iteration 8000 samples were generated.*

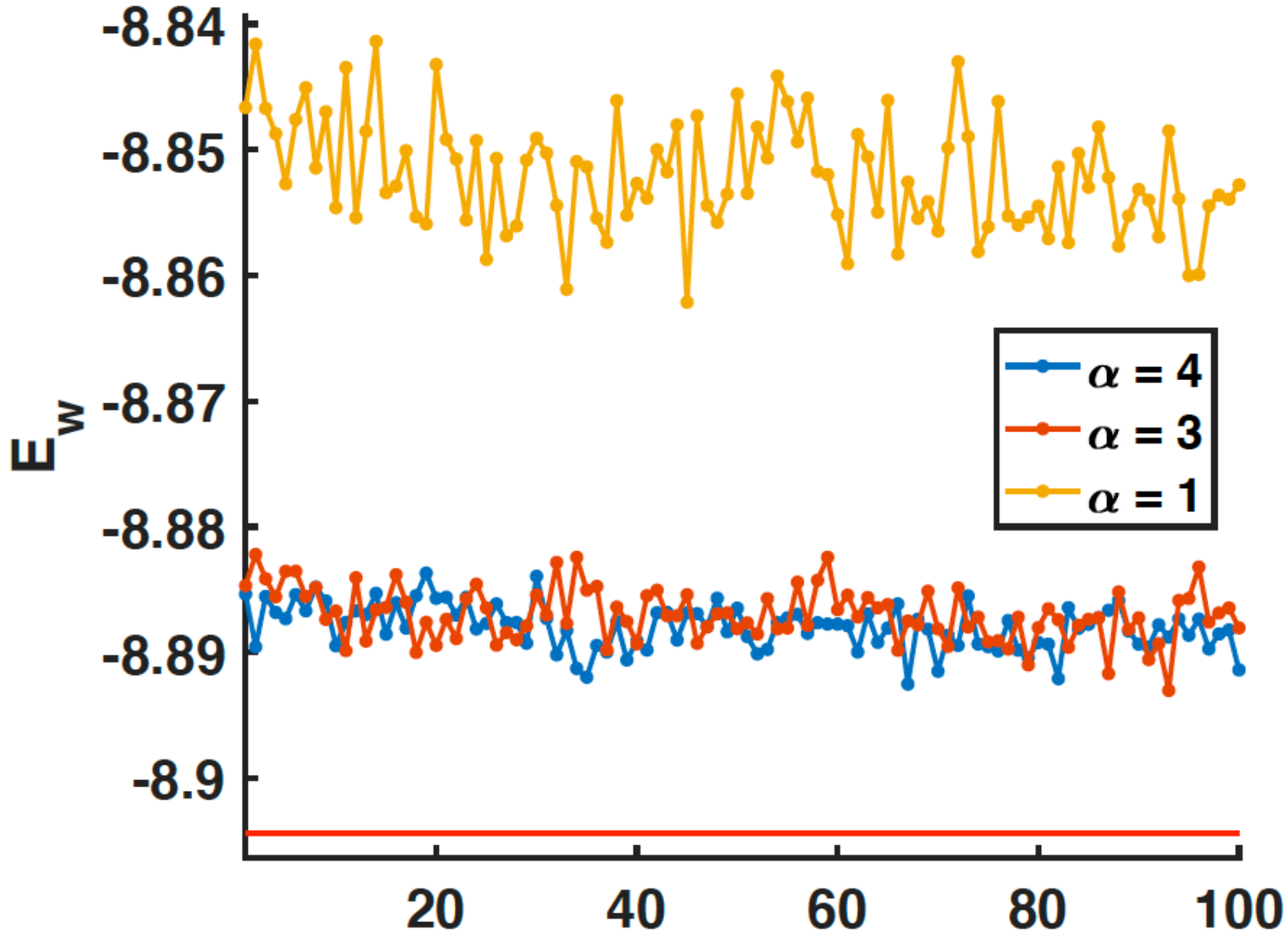



*Figure 7: Energy expectation value at the last* 100 *optimization steps for* $N = 20$ *sites with PBC,* $J = 1$ *and* $\Delta = 1$*, for RBM states with different densities* $\alpha$ *of hidden variables. The red line still marks the exact value of the ground-state energy. For all curves referring to the last optimization steps,* $\gamma = 0.1$ *was kept constant, and at each iteration* 6000 *samples were generated.*