Machine Learning for Many-body Potentials with Moment Tensors

Tobias Olbrich
tobias.olbrich@scai.fraunhofer.de

Fraunhofer Institute for Algorithms and Scientific Computing SCAI
Schloss Birlinghoven, 53754 Sankt Augustin, Germany

October 25, 2018

Outline

Setting

- Given a system of atoms with Cartesian coordinates $\{x_1, \ldots, x_n\} \subset \mathbb{R}^3$ with atomic numbers $\{Z_1, \ldots, Z_n\}$, find the system's energy and its derivative wrt. the atoms' positions.

Setting

- Given a system of atoms with Cartesian coordinates $\{x_1, \ldots, x_n\} \subset \mathbb{R}^3$ with atomic numbers $\{Z_1, \ldots, Z_n\}$, find the system's energy and its derivative wrt. the atoms' positions.
- Mesoscale systems (thousands to hundreds of thousands of atoms).

Classical Approaches

- *Ab-initio* methods: generally too slow
- Empirical potentials, like Morse below: have trouble reaching the desired accuracy

$$E(\{x_i\}) = \sum_{i<j} D_e \left[\exp(-2a(r_{ij} - r_e)) - 2\exp(-a(r_{ij} - r_e))\right]$$

Machine Learning to the Rescue

Key ingredient for ML potentials: invariant descriptors of atomic environments.

[1] Albert Bartók (2010). *The Gaussian Approximation Potential: An Interatomic Potential Derived from First Principles Quantum Mechanics*. Springer Science & Business Media

[2] Jörg Behler (2011). "Atom-centered symmetry functions for constructing high-dimensional neural network potentials". In: *The Journal of chemical physics* 134.7, p. 074106

Machine Learning to the Rescue

Key ingredient for ML potentials: invariant descriptors of atomic environments.

- Gaussian approximation potentials[1] (GAP), using spherical harmonics and Gaussian kernels.
- Neural network potentials[2] (NNP), using a set of *atom-centered symmetry functions* and feed-forward neural networks.

---

[1] Albert Bartók (2010). *The Gaussian Approximation Potential: An Interatomic Potential Derived from First Principles Quantum Mechanics*. Springer Science & Business Media

[2] Jörg Behler (2011). "Atom-centered symmetry functions for constructing high-dimensional neural network potentials". In: *The Journal of chemical physics* 134.7, p. 074106

Machine Learning to the Rescue

Key ingredient for ML potentials: invariant descriptors of atomic environments.

- Gaussian approximation potentials[1] (GAP), using spherical harmonics and Gaussian kernels.
- Neural network potentials[2] (NNP), using a set of *atom-centered symmetry functions* and feed-forward neural networks.

Systematically improvable (given enough training data), but still slow.

---

[1] Albert Bartók (2010). *The Gaussian Approximation Potential: An Interatomic Potential Derived from First Principles Quantum Mechanics*. Springer Science & Business Media

[2] Jörg Behler (2011). "Atom-centered symmetry functions for constructing high-dimensional neural network potentials". In: *The Journal of chemical physics* 134.7, p. 074106

Outline

Basic Idea

- Find invariant polynomial $p(u)$ that approximates the site energy $V(u)$.
- For that, construct set of basis polynomials $b_i \in \mathbb{P}_{perm} \cap \mathbb{P}_{rot}$.
- Then, learn coefficients c_i and set

$$p(u) := \sum_i c_i b_i(u).$$

Alexander V Shapeev (2016). "Moment tensor potentials: A class of systematically improvable interatomic potentials". In: *Multiscale Modeling & Simulation* 14.3, pp. 1153–1173

Moment Tensors

For $\mu, \nu \in \mathbb{N}$, define the moment tensors by

$$M_{\mu,\nu}(u) := \sum_i \|u_i\|^{2\mu} u_i^{\otimes\nu},$$

where

$$u_i^{\otimes\nu} = \underbrace{u_i \otimes \ldots \otimes u_i}_{\nu \text{ times}}$$

is the ν-fold Kronecker product.

The Basis Polynomials

- The final basis functions B_α are constructed as invariant polynomials of the entries of the moment tensors
- Efficient evaluation is possible by using symmetry of the tensors and common sub-terms

Approximation Error Estimate

Theorem (Shapeev)

Let V^q the site-energy of a tight-binding quantum model and $n \leq n_0$. Then there exists $C > 0$ and $\rho > 1$, both depending only on n_0, V^q and the temperature T, such that for any $m \in \mathbb{N}$ there exists $p_m \in \mathbb{P}_{rot} \cap \mathbb{P}_{perm}$ of degree m such that

$$\sup_{u:\, \max_i \|u_i\| < R} |V^q(u) - p_m(u)| < C\rho^{-m}.$$

Representability

Theorem (Shapeev)

The polynomials B_α form a spanning set of the linear space $\mathbb{P}_{rot} \cap \mathbb{P}_{perm} \subset \mathbb{P}$.

Outline

The Cutoff Function

To obtain smoothness at some $R_{cut} > 0$, we define

$$\hat{f}_{\mu,\nu}(r) := \begin{cases} r^{-\nu-2+\mu}(R_{cut} - r)^2 & r < r_{cut} \\ 0 & r \geq R_{cut} \end{cases}$$

and define the function f either by ortho-normalizing the $\hat{f}$ on the interval $[R_{min}, R_{cut}]$ with some weight or by setting

$$f_{\mu,\nu}(r) := C_\mu \hat{f}_{\mu,\nu}(r)$$

with C_μ being the μ-th Chebyshev polynomial. Then

$$M_{\mu,\nu}(u) := \sum_{i=1}^{n} f_{\mu,\nu}(\|u_i\|) u_i^{\otimes\nu}$$

yields the desired property.

Training the Model

For some fixed set A of α and a training set containing configurations indexed by $k = 1, \ldots,$, the least squares system reads

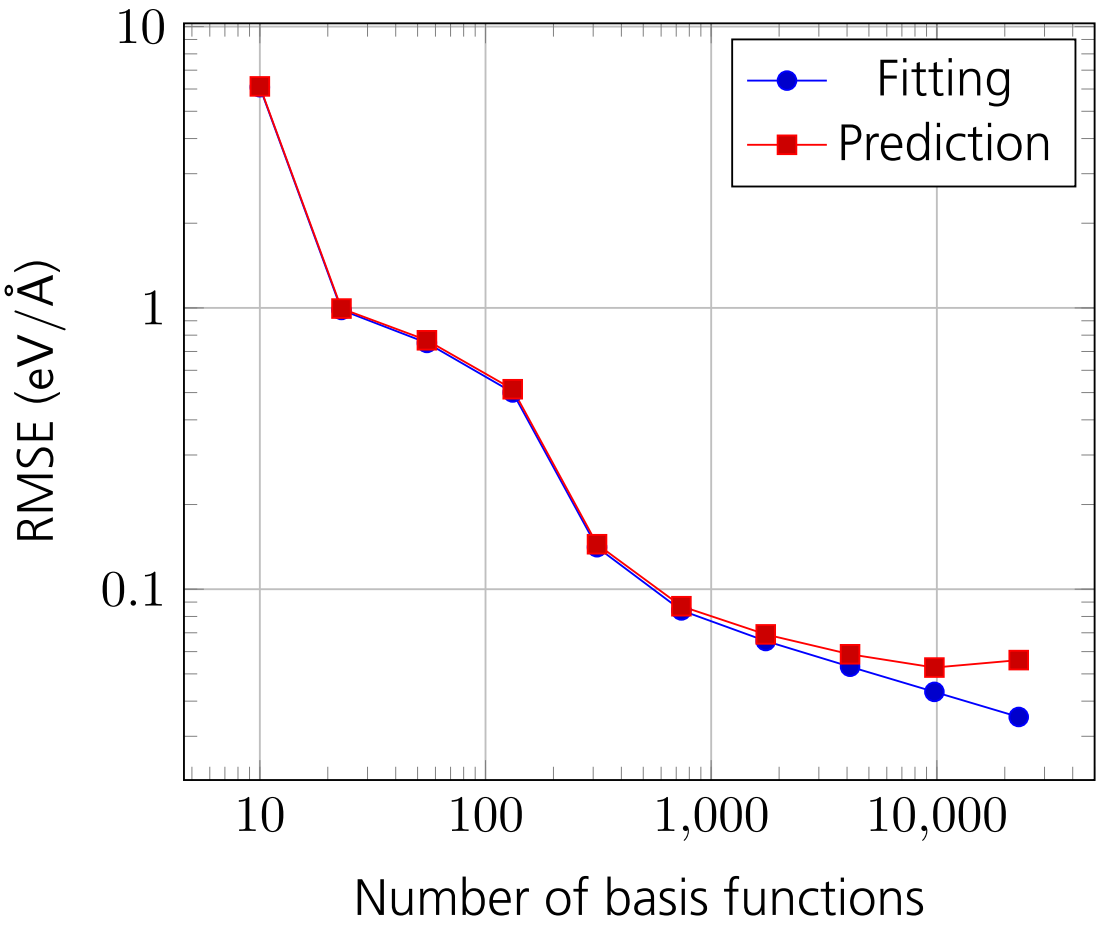
$$\sum_{i=1}^{N^{(k)}} \sum_{\alpha \in A} c_\alpha B_\alpha(Dx_i^{(k)}) = E^{(k)}$$

$$\frac{\partial}{\partial x_j^{(k)}} \sum_{i=1}^{N^{(k)}} \sum_{\alpha \in A} c_\alpha B_\alpha(Dx_i^{(k)}) = -f_j^{(k)},$$

in matrix form with regularization

$$\min_c \|Mc - g\|_{l_2} + \gamma \|c\|_{l_2}.$$

Fitting and Prediction Error for Tungsten Dataset[3]



[3] http://www.libatoms.org/Home/TungstenTrainingConfigurations

Comparison to Other Potentials

Potential	RMSE[4] (eV/Å)
Morse	0.500
Tersoff	0.387
Sti.-We.	0.391
GAP	0.063
MTP	0.051

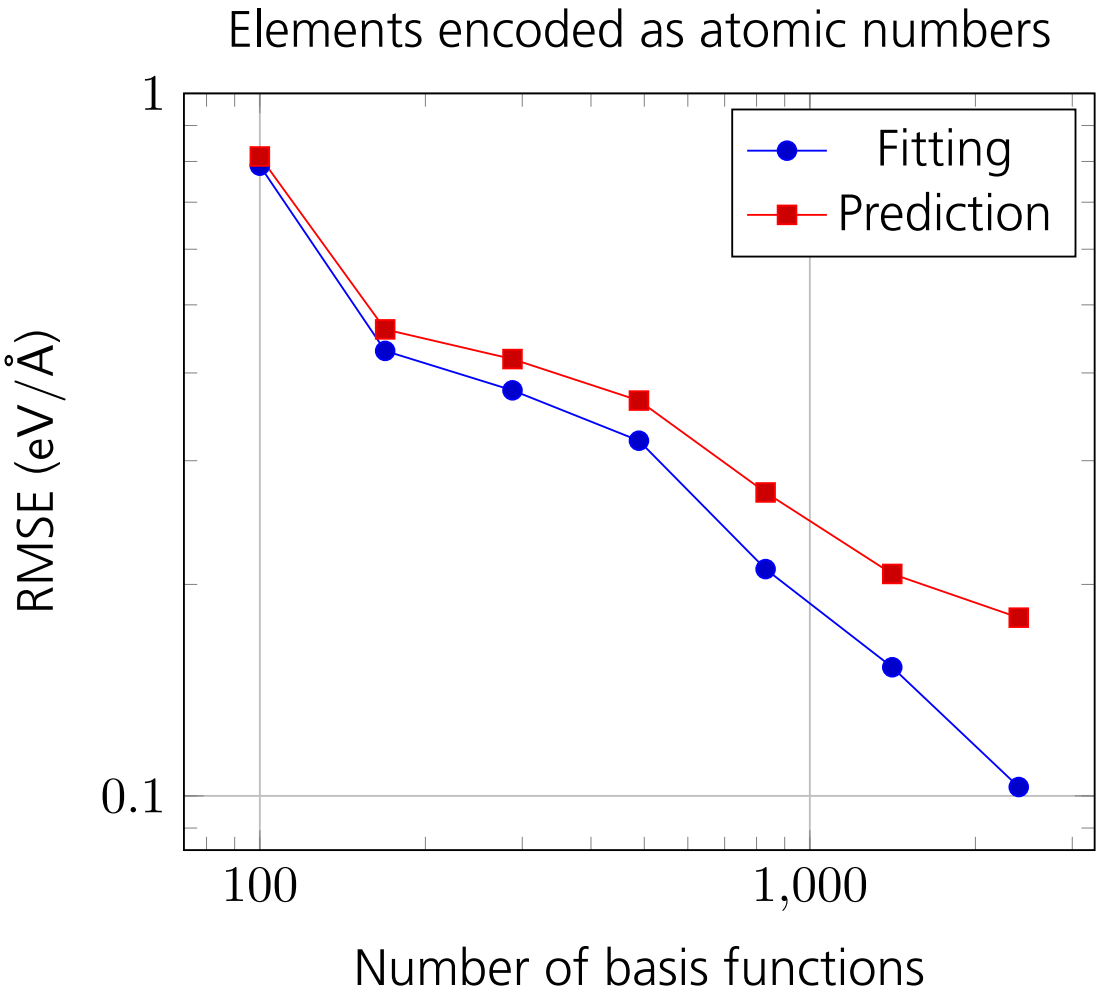
[4] Data for MTP from my experiments, rest of the data by Richard Palme, Fraunhofer SCAI-VMD

The MTP for Systems of Several Particle Types

Idea: Encode the particle types via the cutoff function, for example using their atomic number Z:

$$f_{\mu,\nu}^{(m)}(r_{ij};i,j) := Z_i Z_j f_{\mu,\nu}(r_{ij})$$

Fitting and Prediction Error for Multi-Element Dataset

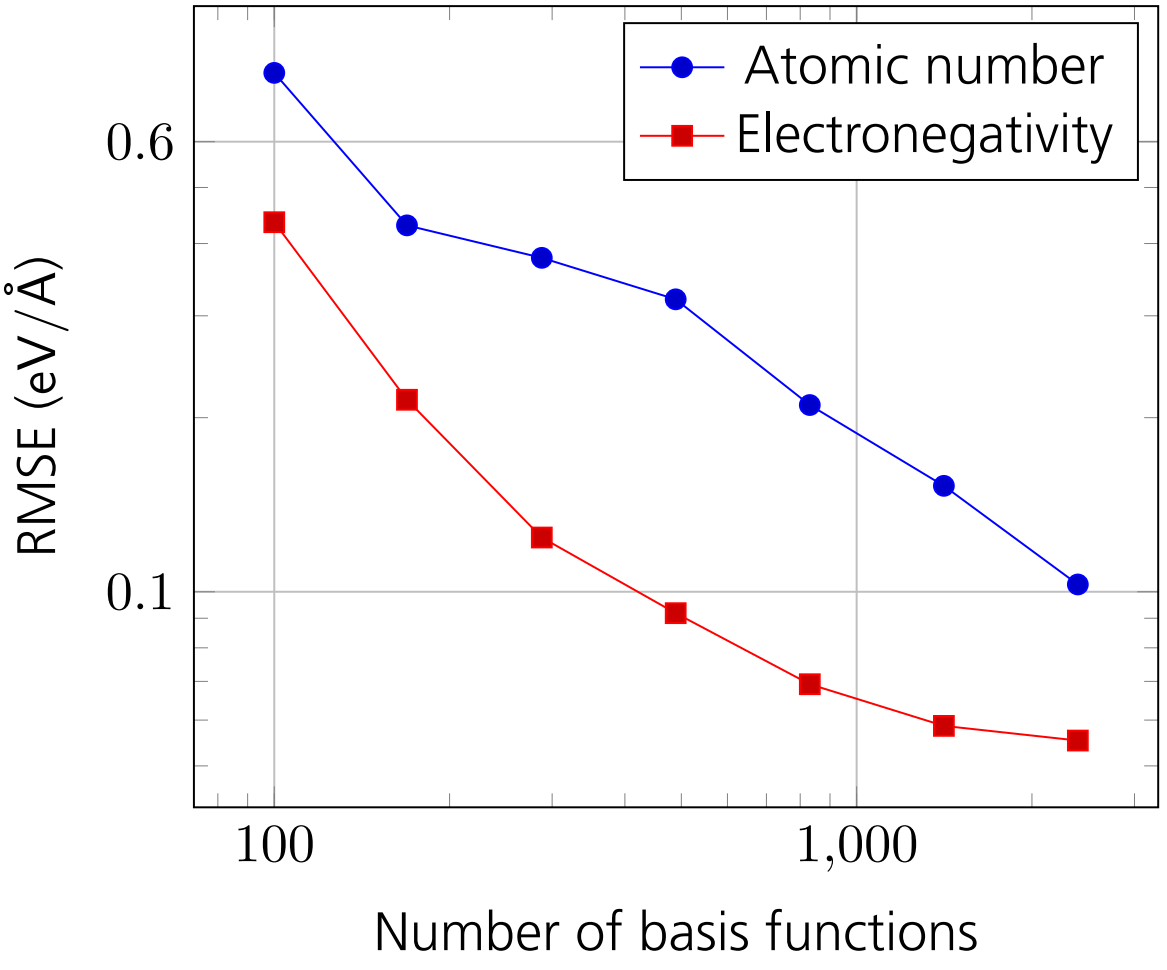


Alternative Encoding of the Particle Types

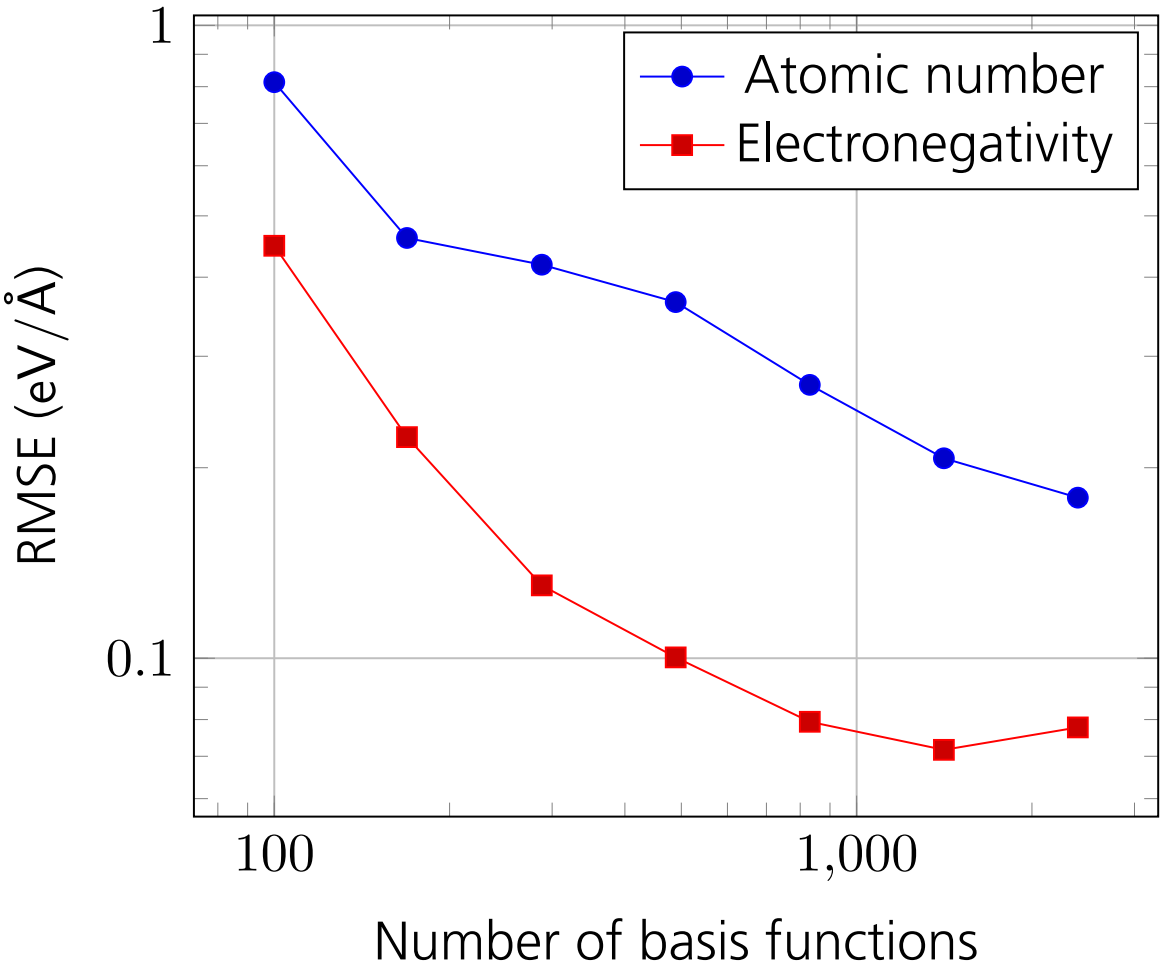
Idea: Using a "more physical" quantity, like the electronegativity χ instead of the atomic numbers:

$$f_{\mu,\nu}^{(m)}(r_{ij};i,j) := \chi_i \chi_j f_{\mu,\nu}(r_{ij})$$

Comparison of Fitting Errors

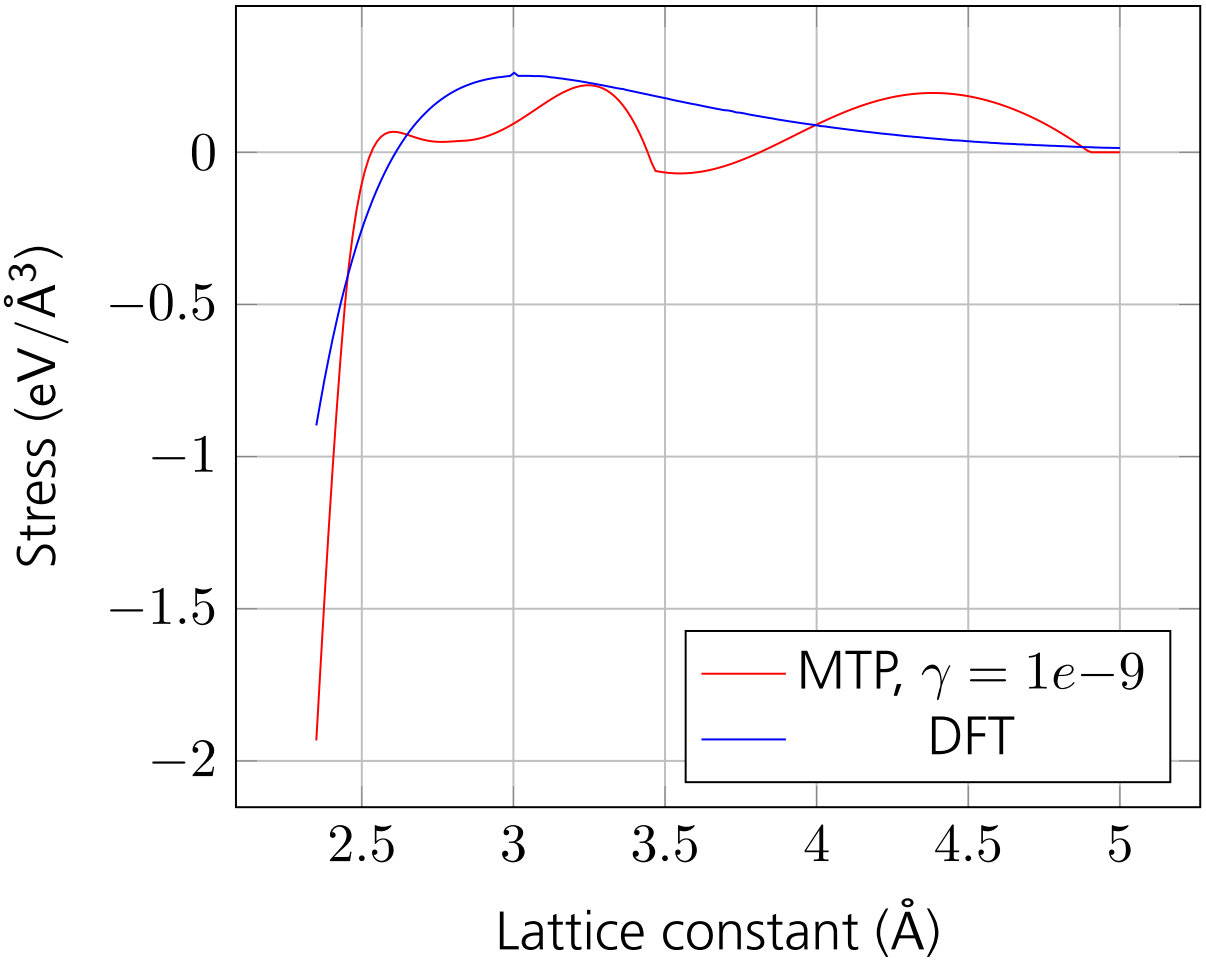


Comparison of Prediction Errors

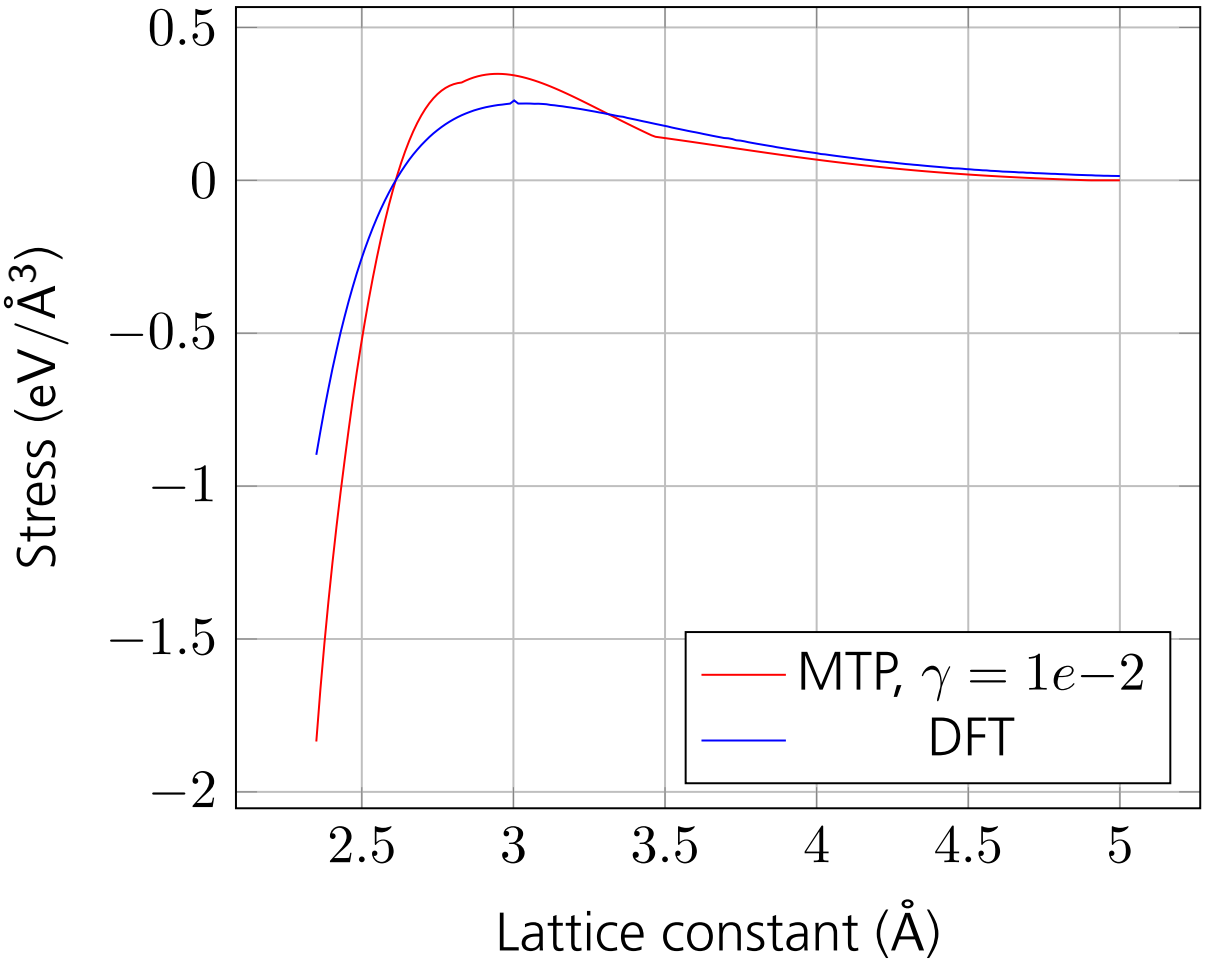


Outline

Overfitting



Stronger Regularization



Fitting Errors Revisited

Potential	RMSE[5] (eV/Å)
Morse	0.500
Sti.-We.	0.391
MTP, $\gamma = 1e{-}9$	0.051
MTP, $\gamma = 1e{-}3$	0.068
MTP, $\gamma = 1e{-}2$	0.075
MTP, $\gamma = 1e{-}1$	0.084

[5] Data for MTP from my experiments, rest of the data by Richard Palme, Fraunhofer SCAI-VMD

Outlook

- Find a better way to pick the hyper-parameter γ

[6] Evgeny V Podryabinkin and Alexander V Shapeev (2017). "Active learning of linearly parametrized interatomic potentials". In: *Computational Materials Science* 140, pp. 171–180

Outlook

- Find a better way to pick the hyper-parameter γ
- Improve smoothness by using other norms for regularization

[6] Evgeny V Podryabinkin and Alexander V Shapeev (2017). "Active learning of linearly parametrized interatomic potentials". In: *Computational Materials Science* 140, pp. 171–180

Outlook

- Find a better way to pick the hyper-parameter γ
- Improve smoothness by using other norms for regularization
- Consider empirical repulsive term, e.g. from Morse potential

[6] Evgeny V Podryabinkin and Alexander V Shapeev (2017). "Active learning of linearly parametrized interatomic potentials". In: *Computational Materials Science* 140, pp. 171–180

Outlook

- Find a better way to pick the hyper-parameter γ
- Improve smoothness by using other norms for regularization
- Consider empirical repulsive term, e.g. from Morse potential
- Reduce the amount of required training data with active learning (based on recent research[6])

[6] Evgeny V Podryabinkin and Alexander V Shapeev (2017). "Active learning of linearly parametrized interatomic potentials". In: *Computational Materials Science* 140, pp. 171–180

Outlook

- Find a better way to pick the hyper-parameter γ
- Improve smoothness by using other norms for regularization
- Consider empirical repulsive term, e.g. from Morse potential
- Reduce the amount of required training data with active learning (based on recent research[6])
- Consider using the invariant polynomials with other ML techniques, such as neural networks

[6] Evgeny V Podryabinkin and Alexander V Shapeev (2017). "Active learning of linearly parametrized interatomic potentials". In: *Computational Materials Science* 140, pp. 171–180

Thank you for your attention!

Questions?