UNIVERSITY OF LIVERPOOL

LEVERHULME TRUST

The Leverhulme Research Centre for Functional Materials Design

Geometry-Informed Machine Learning for Materials Science

Winter School on Robotics and AI for Materials Discovery

Miloslav Torda

3 December 2024

CCDC diamond Hartree Centre Science & Technology Facilities Council Imperial College London ICNS King Abdullah University of Science and Technology UNIVERSITY OF Southampton

Days 1 - 10
Teach yourself variables, constants, arrays, strings, expressions, statements, functions,...

Days 11 - 21
Teach yourself program flow, pointers, references, classes, objects, inheritance, polymorphism, ...

Days 22 - 697
Do a lot of recreational programming. Have fun hacking but remember to learn from your mistakes.

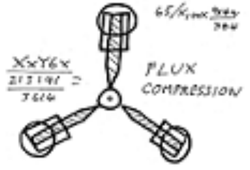
Days 698 - 3648
Interact with other programmers. Work on programming projects together. Learn from them.

Days 3649 - 7781
Teach yourself advanced theoretical physics and formulate a consistent theory of quantum gravity.

Days 7782 - 14611
Teach yourself biochemistry, molecular biology, genetics,...

Day 14611
Use knowledge of biology to make an age-reversing potion.

Day 14611
Use knowledge of physics to build flux capacitor and go back in time to day 21.



Day 21
Replace younger self.

As far as I know, this is the easiest way to "Teach Yourself C++ in 21 Days".

Outline

1. Why think about machine learning geometrically?

Outline

1. Why think about machine learning geometrically?

2. Case Study: Molecular Crystal Structure Prediction.

Outline

1. Why think about machine learning geometrically?

2. Case Study: Molecular Crystal Structure Prediction.

3. Closing remarks.

1. Why think about machine learning geometrically?

Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)$ $(i = 1, \ldots, N)$

1. Why think about machine learning geometrically?

Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)\ (i = 1, \ldots, N)$

Linear Algebra

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

1. Why think about machine learning geometrically?

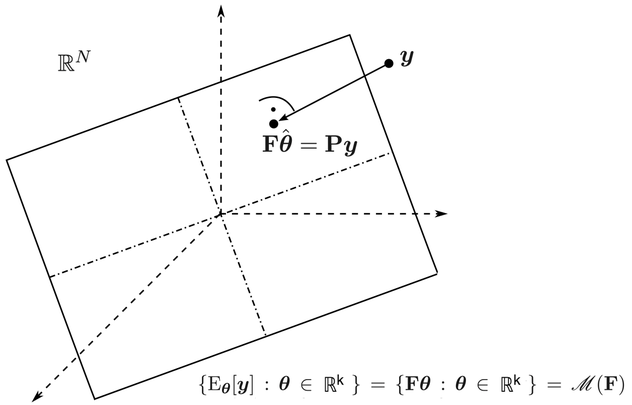
Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)$ $(i = 1, \ldots, N)$

Linear Algebra and Geometry

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i, \; i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$



1. Why think about machine learning geometrically?

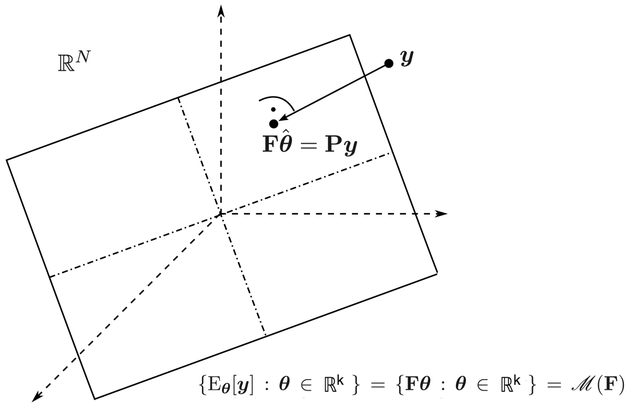
Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)$ $(i = 1, \ldots, N)$

Linear Algebra and Geometry

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$



$\mathscr{E}$

Learning means finding values of θ from measurements

1. Why think about machine learning geometrically?

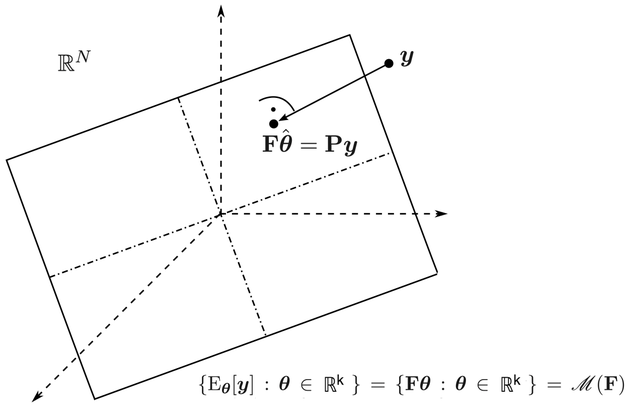
Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)\ (i = 1, \ldots, N)$

Linear Algebra and Geometry

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$



Learning means finding values of θ from measurements

Least Squares estimate $\hat{\theta}$ of θ is an orthogonal projection of y onto the plane of expected values $\mathscr{M}(\mathbf{F})$.

1. Why think about machine learning geometrically?

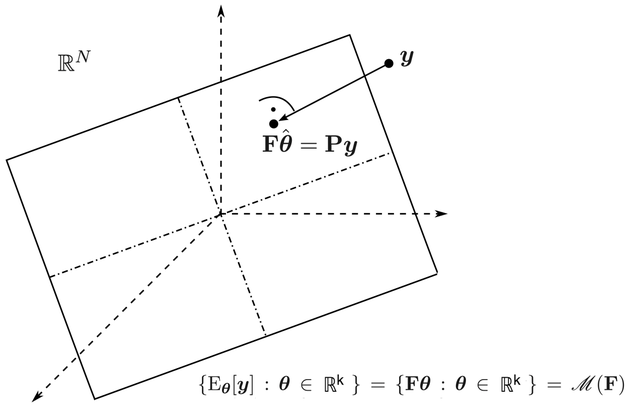
Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)\ (i = 1, \ldots, N)$

Linear Algebra and Geometry

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$



Learning means finding values of θ from measurements

Least Squares estimate $\hat{\theta}$ of θ is an orthogonal projection of y onto the plane of expected values $\mathscr{M}(\mathbf{F})$.

Example. $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \theta_0 + \epsilon$ $\qquad \hat{\theta}_0 = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y_i}$

1. Why think about machine learning geometrically?

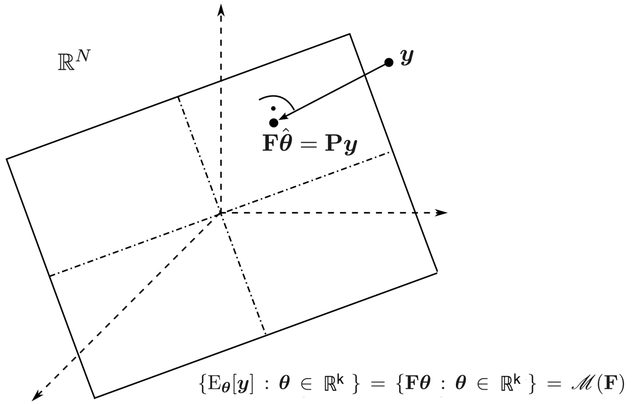
Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)\ (i = 1, \ldots, N)$

Linear Algebra and Geometry

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

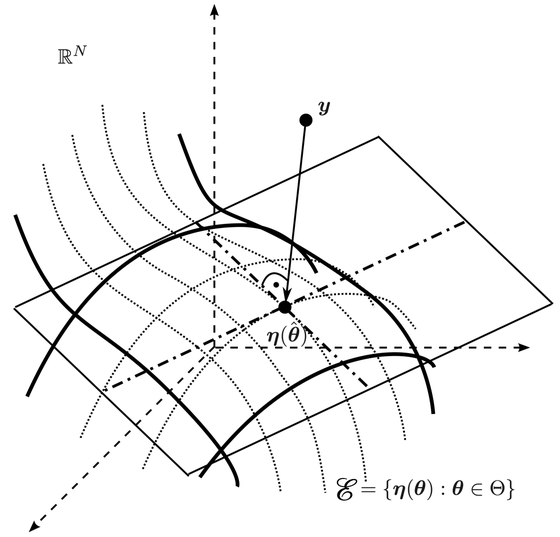


Learning means finding values of θ from measurements

Least Squares estimate $\hat{\theta}$ of θ is an orthogonal projection of y onto the plane of expected values $\mathscr{M}(\mathbf{F})$.

Example. $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \theta_0 + \epsilon$ $\qquad \hat{\theta}_0 = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y_i}$

Bad news. Usually $\mathscr{M}$ is not a plane (linear) but a surface $\mathscr{E}$ or some other space.



1. Why think about machine learning geometrically?

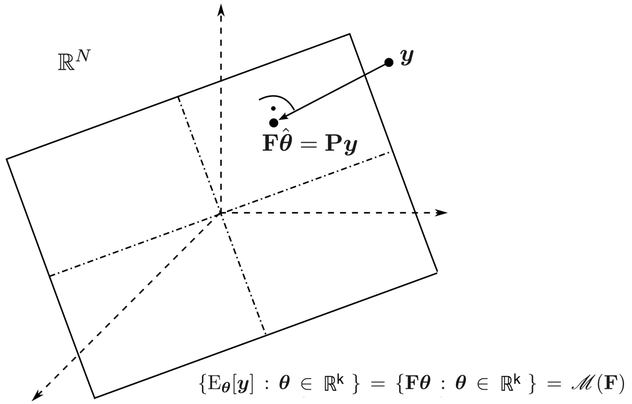
Geometry of the Linear Regression Model

N measurements. $(y_i, x_i^1, \ldots, x_i^k)\ (i = 1, \ldots, N)$

Linear Algebra and Geometry

$$y_i = \theta_0 + \theta_1 x_i + \cdots + \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, N$$

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^k \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

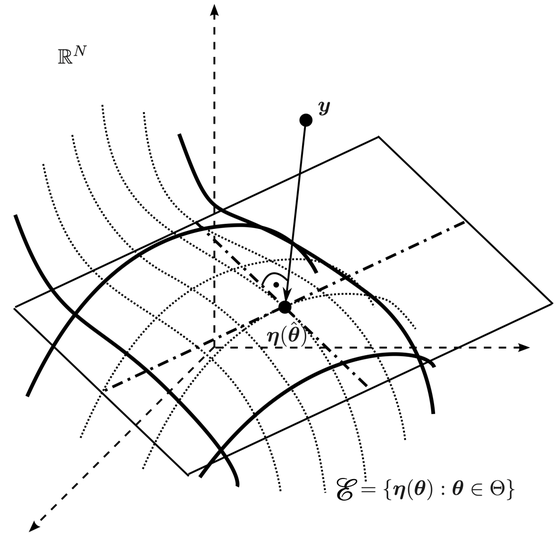


Learning means finding values of θ from measurements

Least Squares estimate $\hat{\theta}$ of θ is an orthogonal projection of y onto the plane of expected values $\mathscr{M}(\mathbf{F})$.

Example. $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \theta_0 + \epsilon$ $\qquad \hat{\theta}_0 = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y_i}$

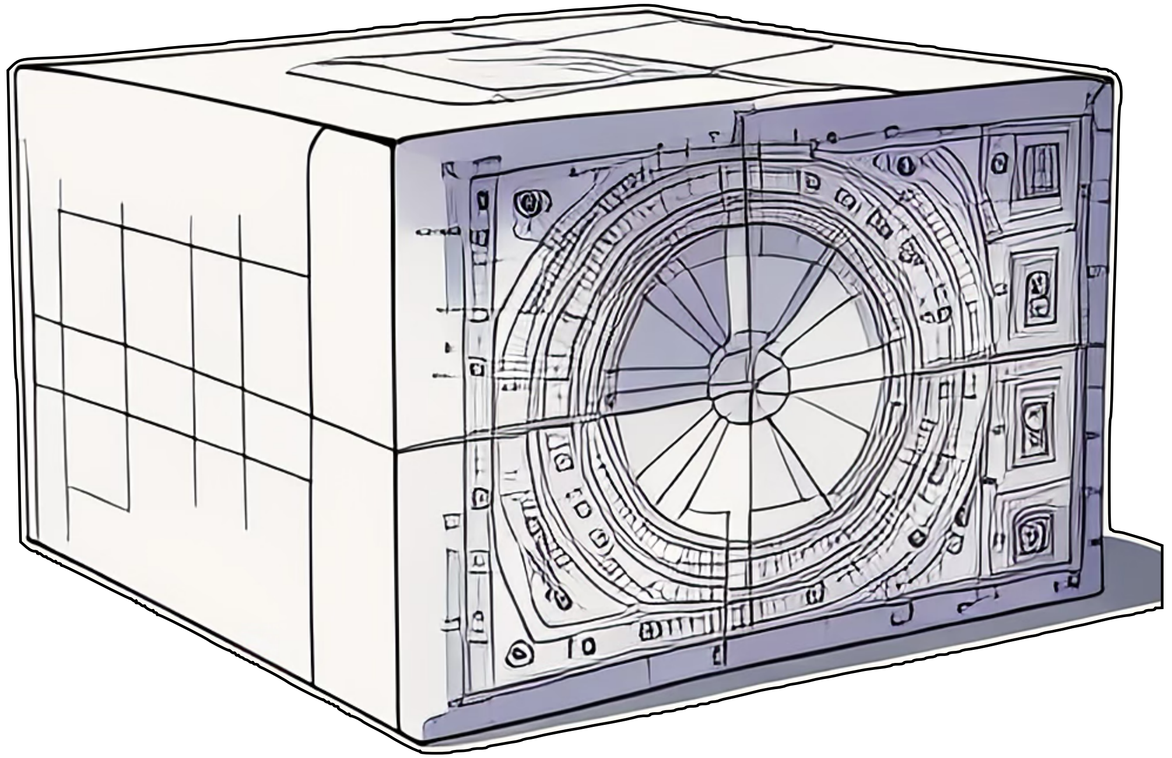
Bad news. Usually $\mathscr{M}$ is not a plane (linear) but a surface $\mathscr{E}$ or some other space.



However, the underlying idea is the same.

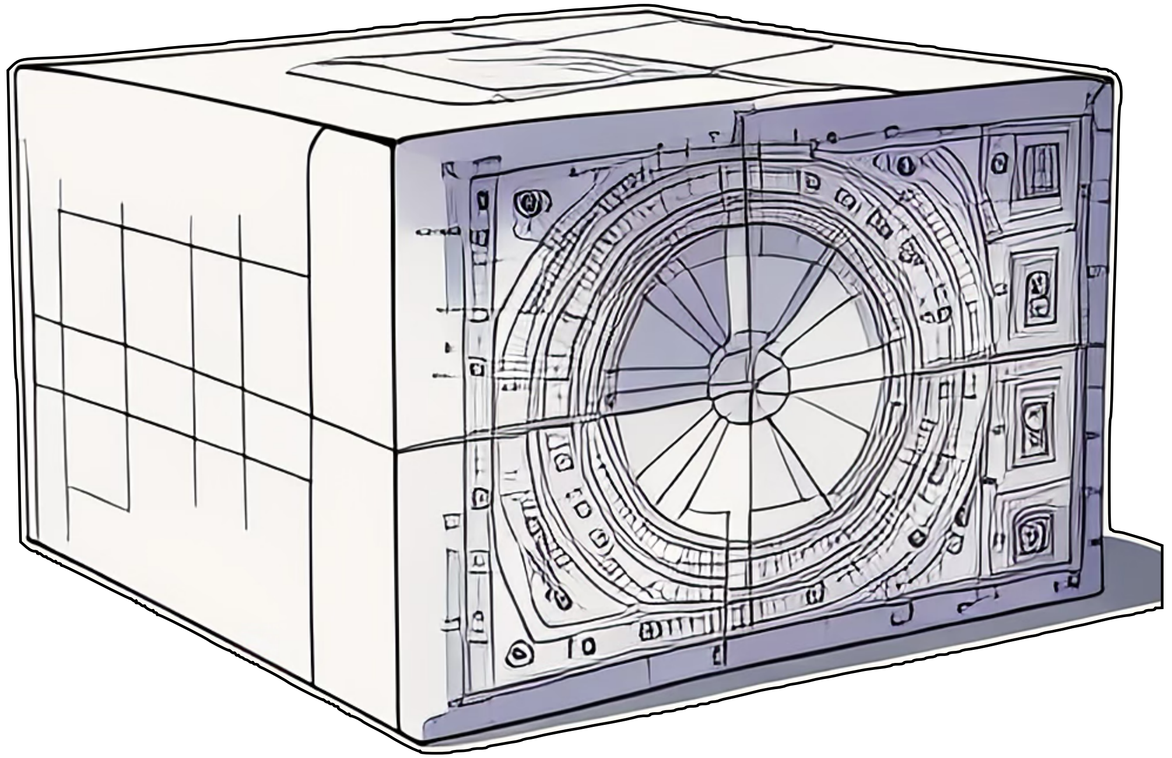
1. Why think about machine learning geometrically?

“All models are wrong, but some are useful.“ - George E. P. Box



1. Why think about machine learning geometrically?

“All models are wrong, but some are useful.“ - George E. P. Box



More complicated models don’t necessarily mean better models.

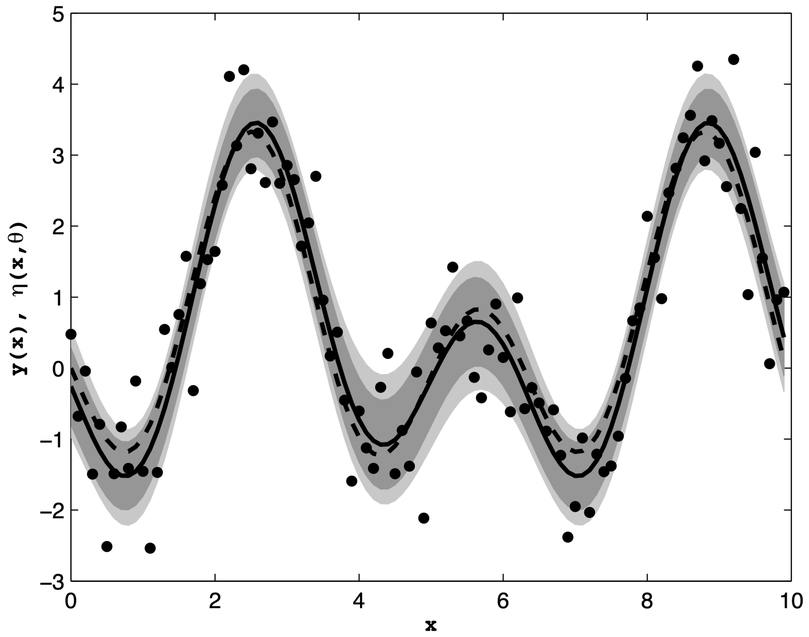
1. Why think about machine learning geometrically?

“All models are wrong, but some are useful.“ - George E. P. Box

Trigonometric regression

Model 1. $y = 0.4514 + 0.6077\sin(x) - 1.2837\cos(x) - 1.4938\sin(2x) + 0.5741\cos(2x)$

Model 2. $y = 0.4521 + 0.6065\sin(x) - 1.2916\cos(x) - 1.4803\sin(2x) + 0.5728\cos(2x) + 0.0155\sin(3x) + 0.0584\cos(3x)$



More complicated models don’t necessarily mean better models.

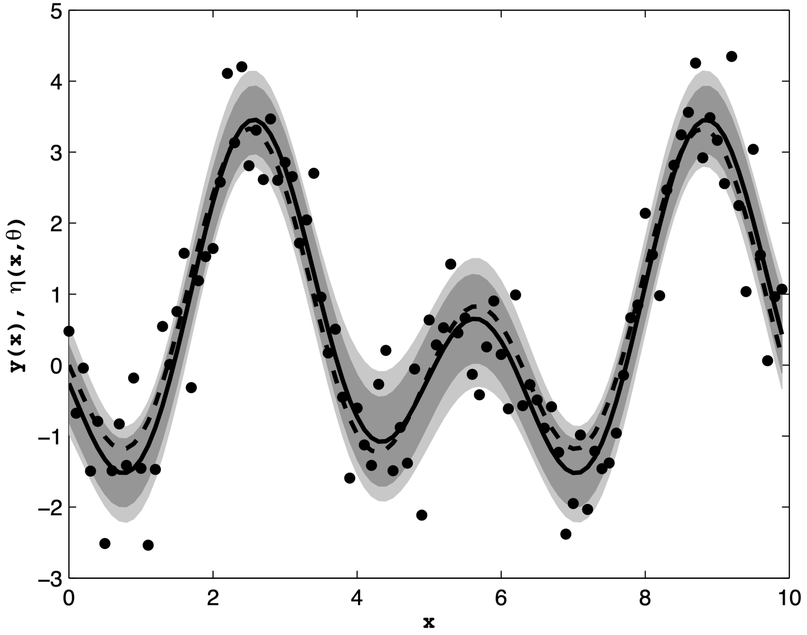
1. Why think about machine learning geometrically?

“All models are wrong, but some are useful.“ - George E. P. Box

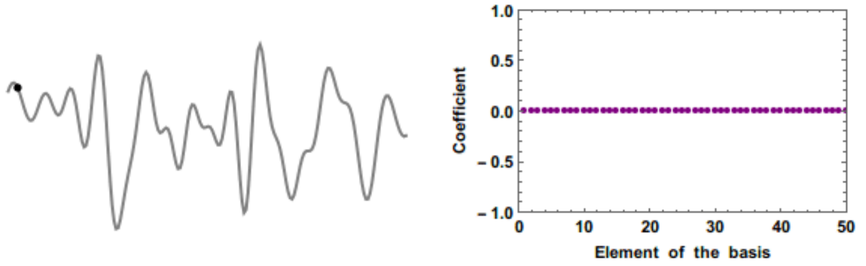
Trigonometric regression

Model 1. $y = 0.4514 + 0.6077\sin(x) - 1.2837\cos(x) - 1.4938\sin(2x) + 0.5741\cos(2x)$

Model 2. $y = 0.4521 + 0.6065\sin(x) - 1.2916\cos(x) - 1.4803\sin(2x) + 0.5728\cos(2x) + 0.0155\sin(3x) + 0.0584\cos(3x)$



Example. Compressed Sensing / Sparse Sampling

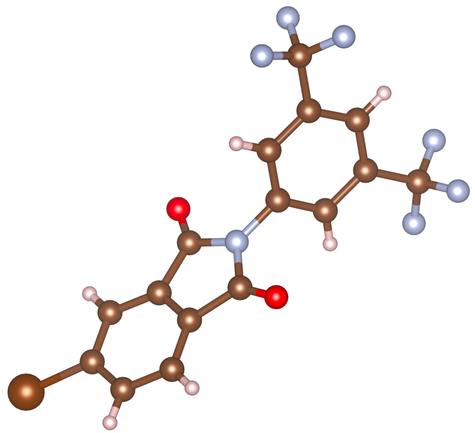


Jacopo Bertolotti, CC0, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Orthogonal_Matching_Pursuit.gif

More complicated models don’t necessarily mean better models.

2. Case Study: Molecular Crystal Structure Prediction.

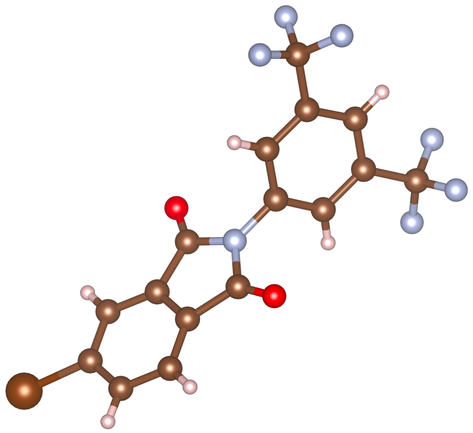
Physical problem. Predict the arrangement of molecules within a crystalline solid based only on the chemical composition of the target molecule from first-principles (i.e. thermodynamics, quantum mechanics)



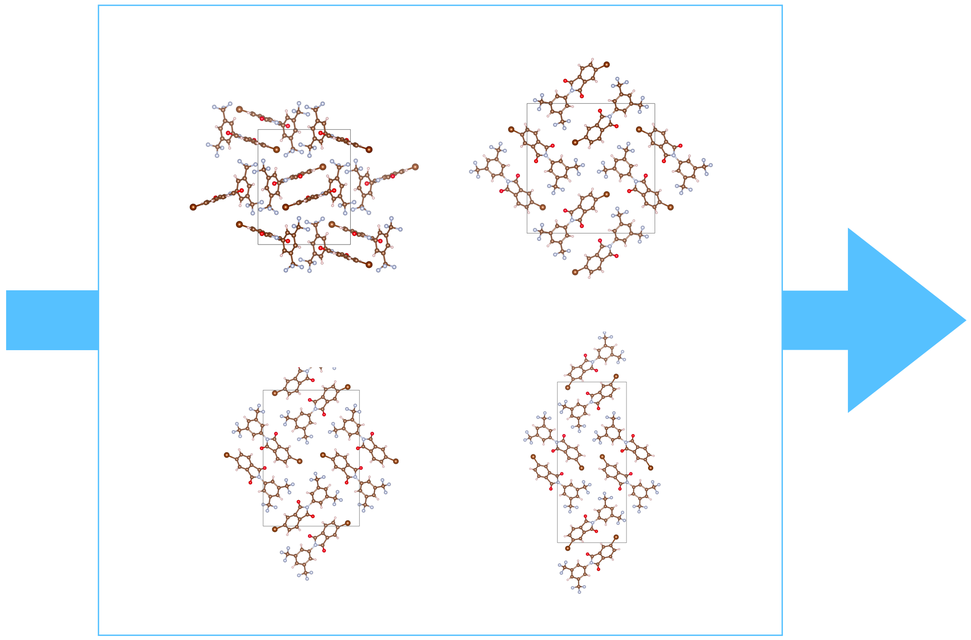
Chemical composition of the target molecule

2. Case Study: Molecular Crystal Structure Prediction.

Physical problem. Predict the arrangement of molecules within a crystalline solid based only on the chemical composition of the target molecule from first-principles (i.e. thermodynamics, quantum mechanics)



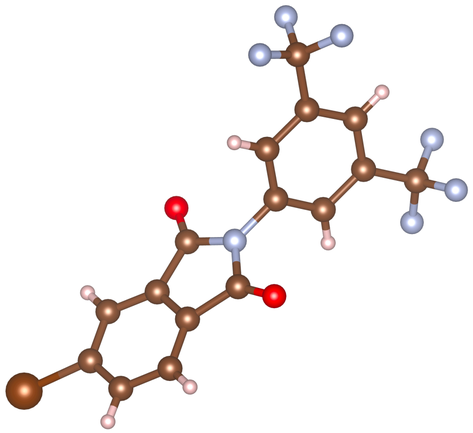
Chemical composition of the target molecule



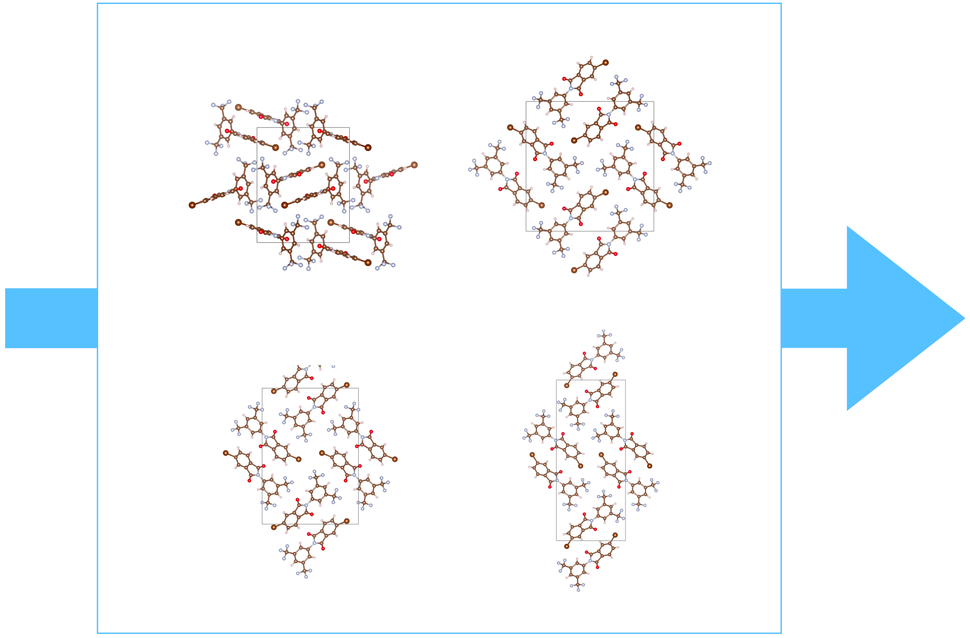
Force-field based structure generation and filtering

2. Case Study: Molecular Crystal Structure Prediction.

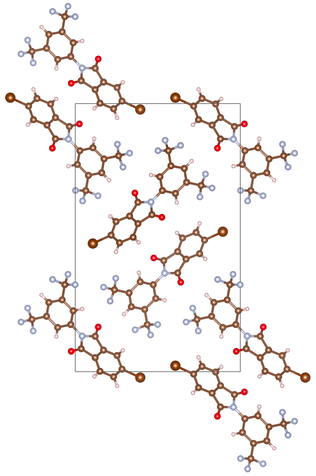
Physical problem. Predict the arrangement of molecules within a crystalline solid based only on the chemical composition of the target molecule from first-principles (i.e. thermodynamics, quantum mechanics)



Chemical composition of the target molecule



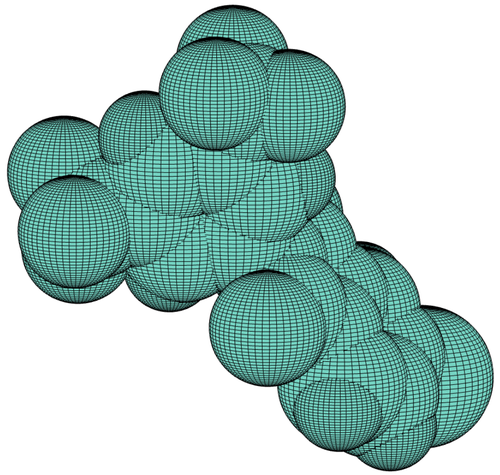
Force-field based structure generation and filtering



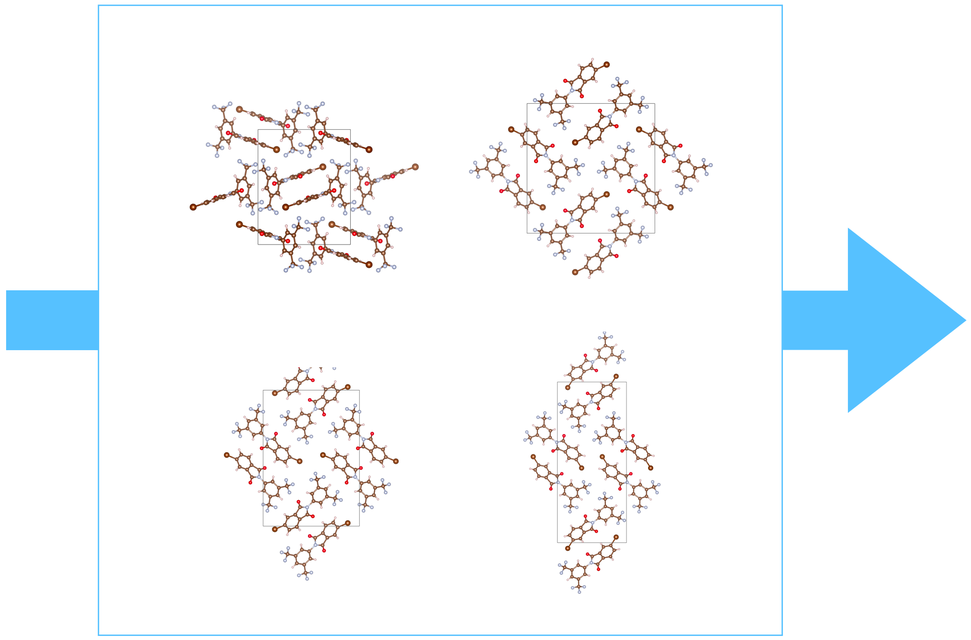
High-Accuracy DFT Geometry Refinement and Final Structure Assessment

2. Case Study: Molecular Crystal Structure Prediction.

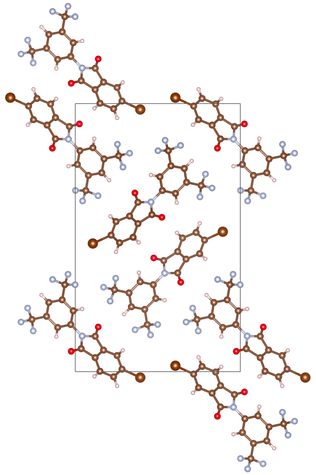
Physical problem. Predict the arrangement of molecules within a crystalline solid based only on the chemical composition of the target molecule from first-principles (i.e. thermodynamics, quantum mechanics)



Geometric Representation of a Molecule



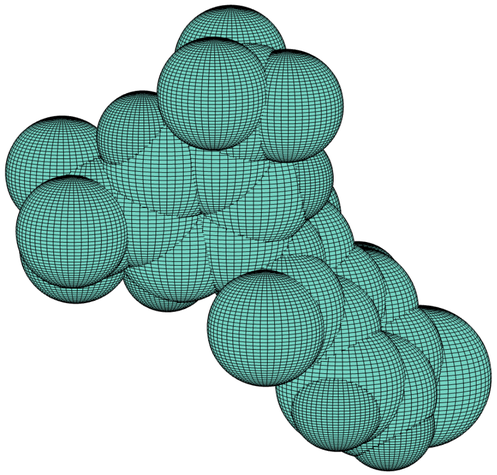
Force-field based structure generation and filtering



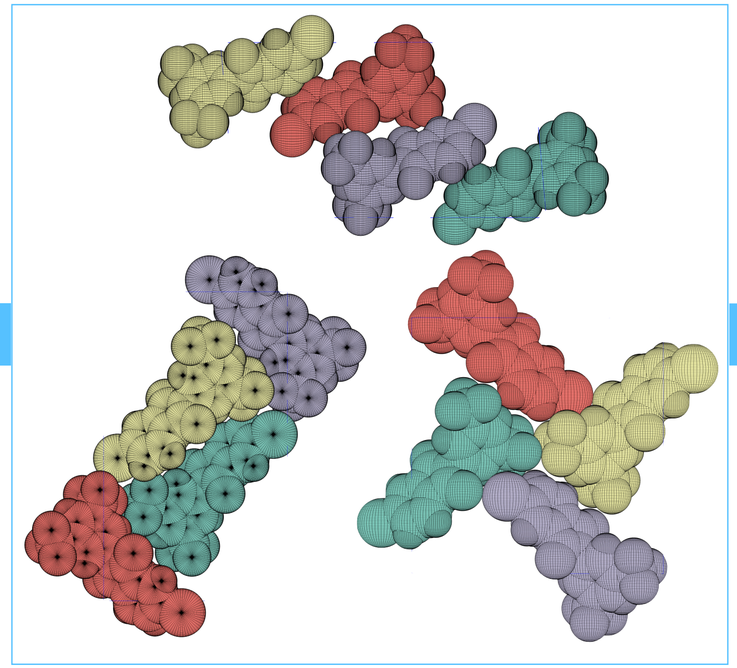
High-Accuracy DFT Geometry Refinement and Final Structure Assessment

2. Case Study: Molecular Crystal Structure Prediction.

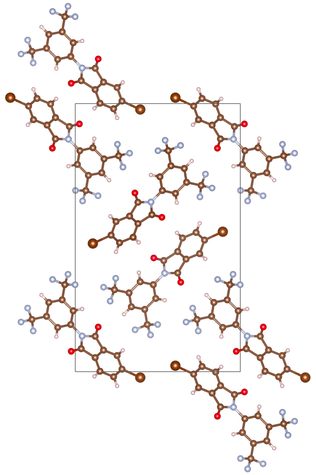
Physical problem. Predict the arrangement of molecules within a crystalline solid based only on the chemical composition of the target molecule from first-principles (i.e. thermodynamics, quantum mechanics)



Geometric Representation of a Molecule



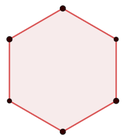
Geometric-Inductive-Biased Structure Generation



High-Accuracy DFT Geometry Refinement and Final Structure Assessment

2. Case Study: Molecular Crystal Structure Prediction.

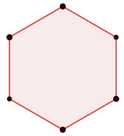
Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

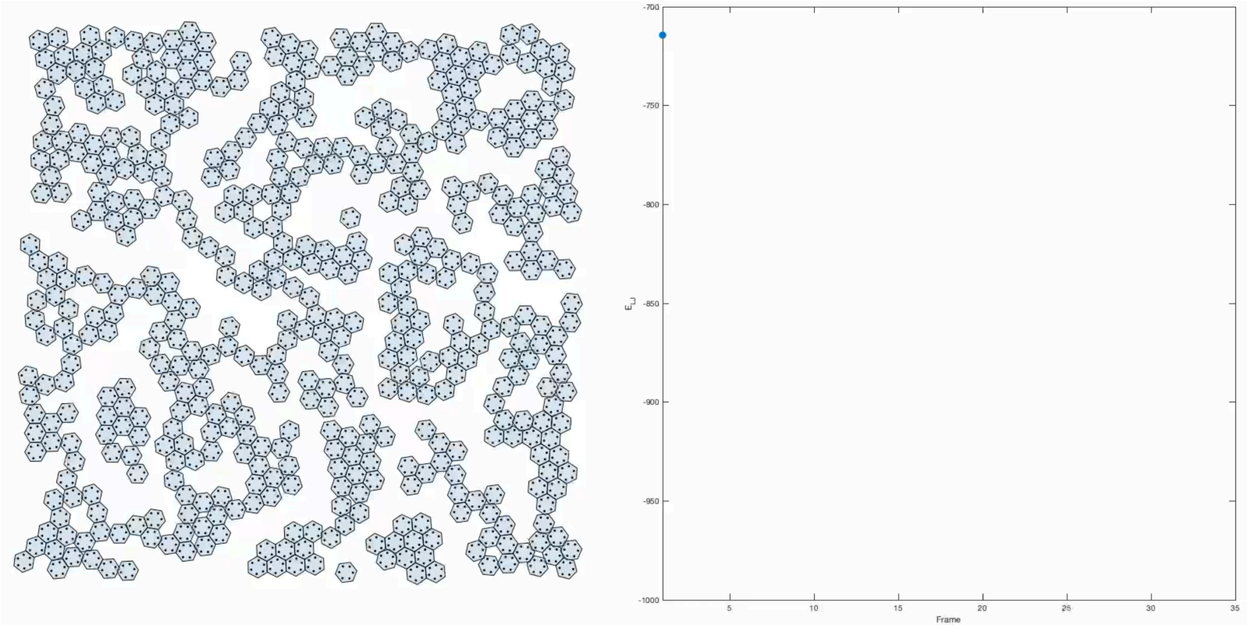


2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.



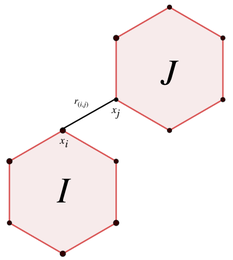


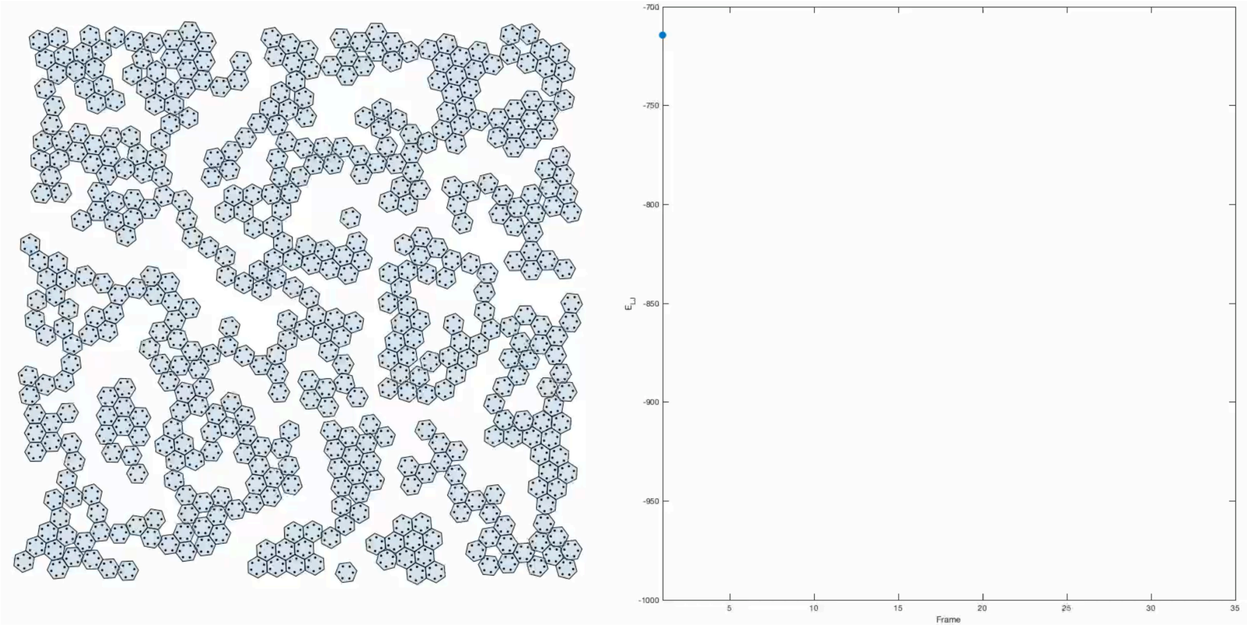
2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.

Intermolecular energy: $E(I,J) = \frac{1}{2}\sum_{i\in I}\sum_{j\in J}\left[\left(\frac{1}{r_{(i,j)}}\right)^{12} - \left(\frac{1}{r_{(i,j)}}\right)^{6}\right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$





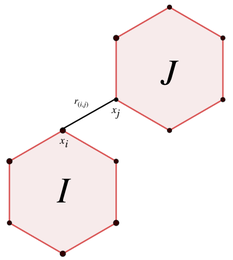
2. Case Study: Molecular Crystal Structure Prediction.

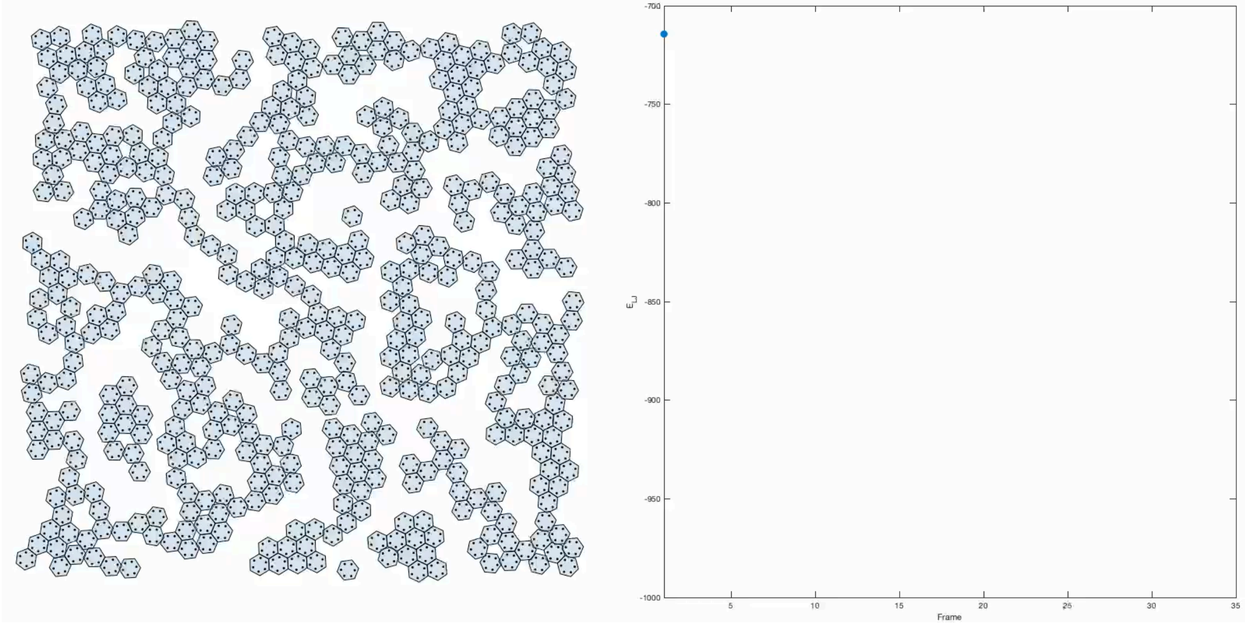
Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.

Intermolecular energy: $E(I,J) = \frac{1}{2} \sum_{i \in I} \sum_{j \in J} \left[\left(\frac{1}{r_{(i,j)}} \right)^{12} - \left(\frac{1}{r_{(i,j)}} \right)^{6} \right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

Ground state configuration given by $\min \sum_{I \cap J = 0} E(I,J) \Rightarrow$ 1500 variables to optimize





2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

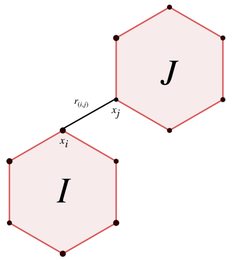
A system of 500 model molecules in a square box.

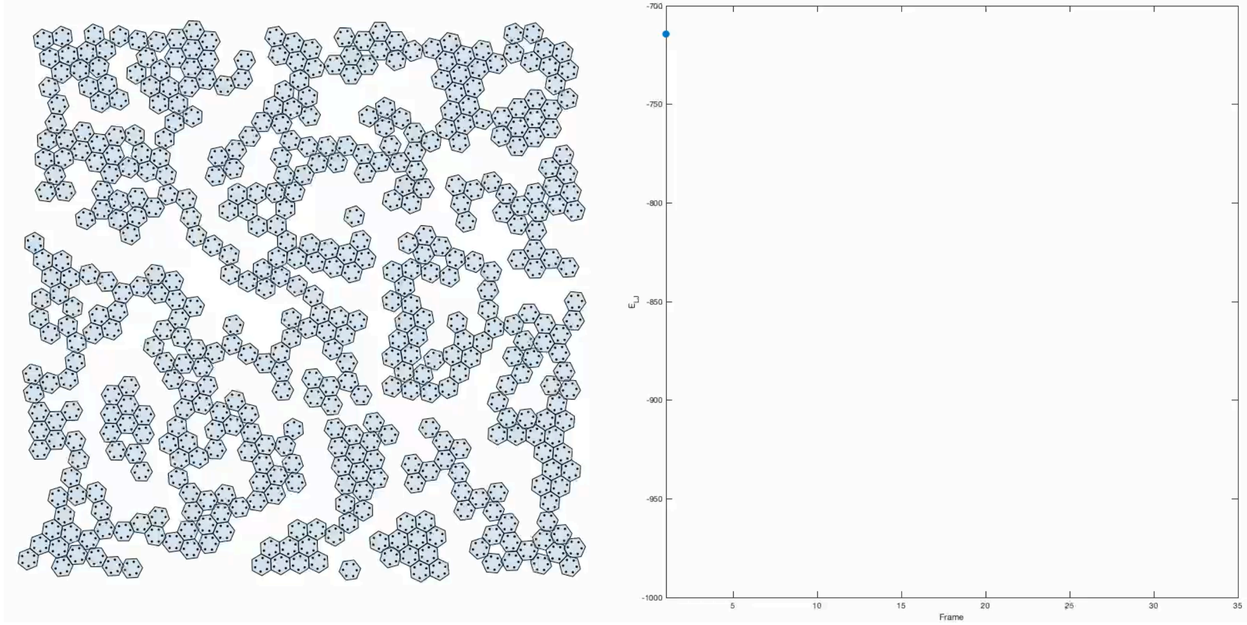
Intermolecular energy: $E(I,J) = \frac{1}{2}\sum_{i\in I}\sum_{j\in J}\left[\left(\frac{1}{r_{(i,j)}}\right)^{12} - \left(\frac{1}{r_{(i,j)}}\right)^{6}\right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

Ground state configuration given by $\min \sum_{I\cap J=0} E(I,J) \Rightarrow$ 1500 variables to optimize

Each frame is an output of a Sequential Quadratic Programming run

- Random initial configuration
- Boundaries are reduced after each iteration





2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

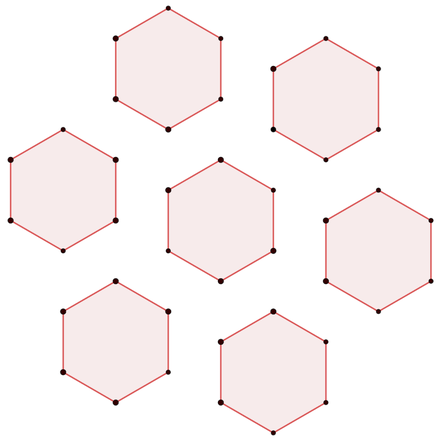
A system of 500 model molecules in a square box.

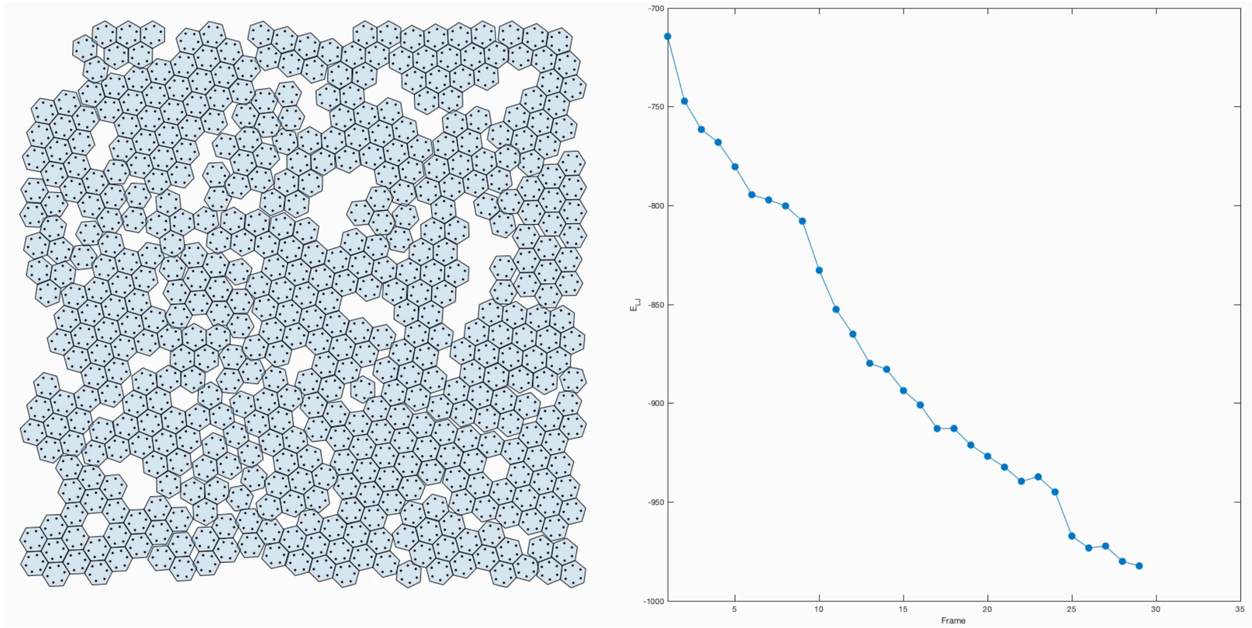
Intermolecular energy: $E(I,J) = \frac{1}{2}\sum_{i\in I}\sum_{j\in J}\left[\left(\frac{1}{r_{(i,j)}}\right)^{12} - \left(\frac{1}{r_{(i,j)}}\right)^{6}\right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

Ground state configuration given by $\min \sum_{I\cap J=0} E(I,J) \Rightarrow$ 1500 variables to optimize

Each frame is an output of a Sequential Quadratic Programming run

- Random initial configuration
- Boundaries are reduced after each iteration





2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

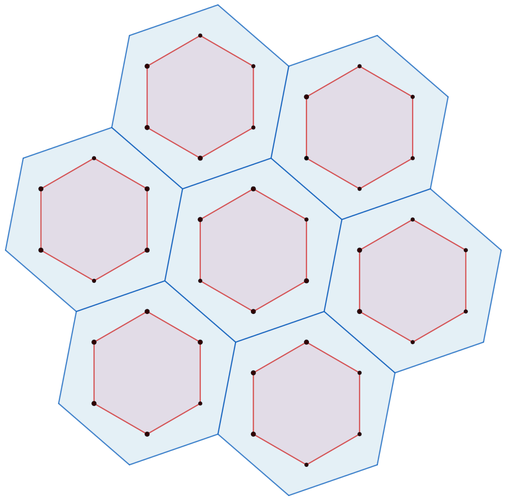
A system of 500 model molecules in a square box.

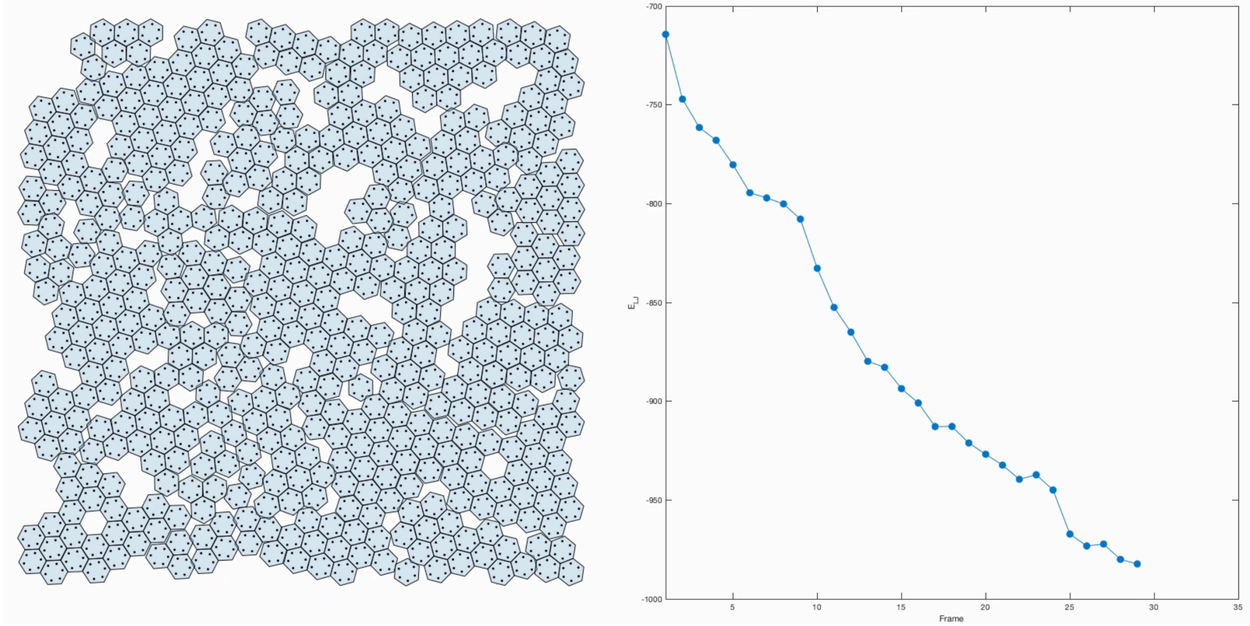
Intermolecular energy: $E(I,J) = \frac{1}{2}\sum_{i\in I}\sum_{j\in J}\left[\left(\frac{1}{r_{(i,j)}}\right)^{12} - \left(\frac{1}{r_{(i,j)}}\right)^{6}\right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

Ground state configuration given by $\min \sum_{I\cap J=0} E(I,J) \Rightarrow$ 1500 variables to optimize

Each frame is an output of a Sequential Quadratic Programming run

- Random initial configuration
- Boundaries are reduced after each iteration





2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.

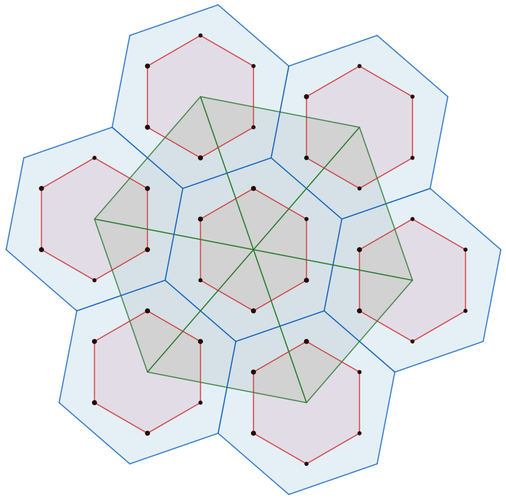
Intermolecular energy: $E(I,J) = \frac{1}{2} \sum_{i \in I} \sum_{j \in J} \left[\left(\frac{1}{r_{(i,j)}} \right)^{12} - \left(\frac{1}{r_{(i,j)}} \right)^{6} \right]$ $r_{(i,j)} = \|x_i - x_j\|$ $x_* \in \mathbb{R}^2$

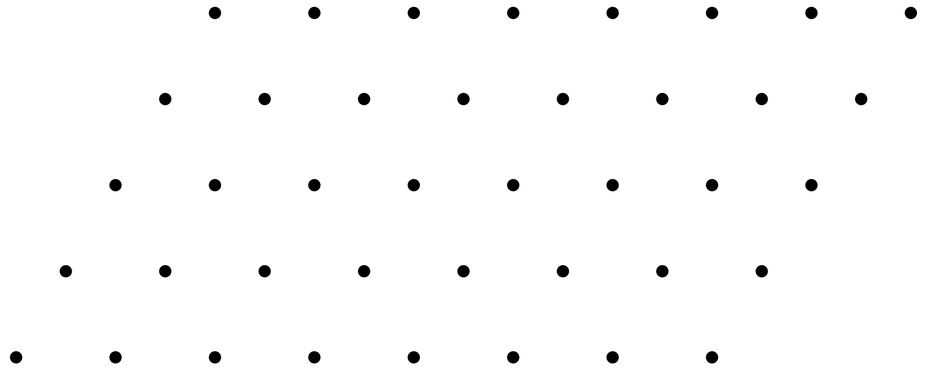
Ground state configuration given by $\min \sum_{I \cap J = 0} E(I,J) \Rightarrow$ 1500 variables to optimize

Each frame is an output of a Sequential Quadratic Programming run

- Random initial configuration
- Boundaries are reduced after each iteration

Ground state configuration is given only by translates of the molecule by a lattice.





2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.

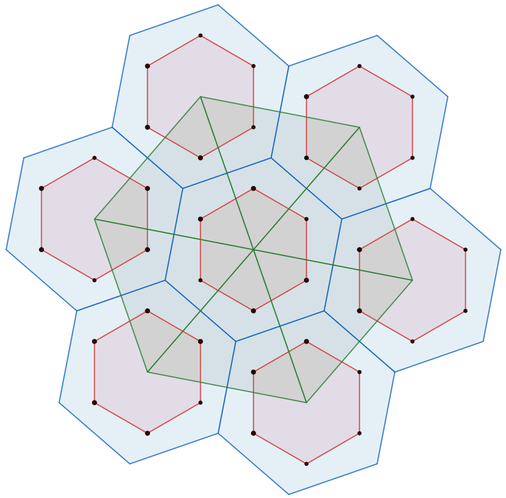
Intermolecular energy: $E(I,J) = \frac{1}{2} \sum_{i \in I} \sum_{j \in J} \left[\left(\frac{1}{r_{(i,j)}} \right)^{12} - \left(\frac{1}{r_{(i,j)}} \right)^{6} \right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

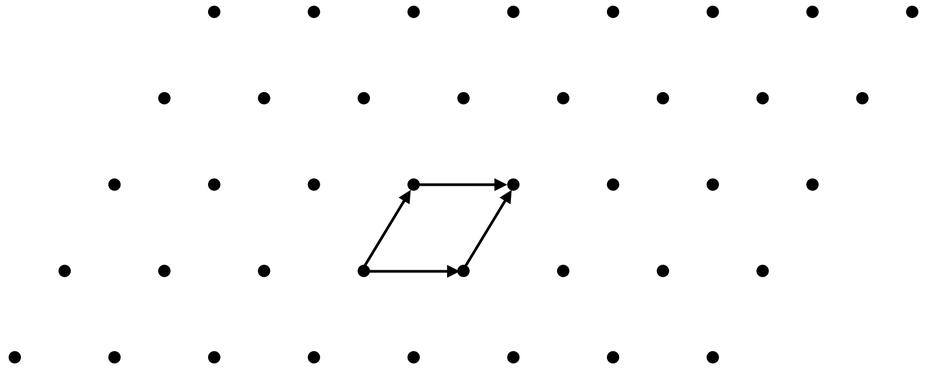
Ground state configuration given by $\min \sum_{I \cap J = 0} E(I,J) \Rightarrow$ 1500 variables to optimize

Each frame is an output of a Sequential Quadratic Programming run

- Random initial configuration
- Boundaries are reduced after each iteration

Ground state configuration is given only by translates of the molecule by a lattice.





2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.

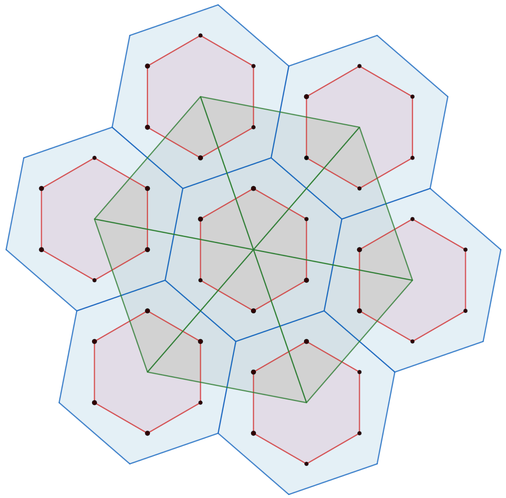
Intermolecular energy: $E(I,J) = \frac{1}{2}\sum_{i\in I}\sum_{j\in J}\left[\left(\frac{1}{r_{(i,j)}}\right)^{12} - \left(\frac{1}{r_{(i,j)}}\right)^{6}\right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

Ground state configuration given by $\min \sum_{I\cap J=0} E(I,J) \Rightarrow$ 1500 variables to optimize

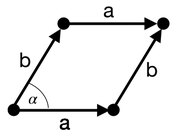
Each frame is an output of a Sequential Quadratic Programming run

- Random initial configuration
- Boundaries are reduced after each iteration

Ground state configuration is given only by translates of the molecule by a lattice.



For ground state prediction it is enough to optimize 3 variables $[a, b, \alpha]$ instead of 1500.



2. Case Study: Molecular Crystal Structure Prediction.

Rigid molecule with 6 general atoms. Configuration of molecule $\xi = \begin{pmatrix} c_1 \\ c_2 \\ \delta \end{pmatrix}$

A system of 500 model molecules in a square box.

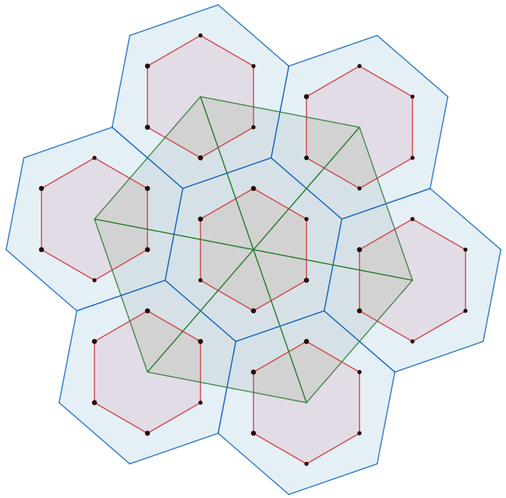
Intermolecular energy: $E(I,J) = \frac{1}{2}\sum_{i\in I}\sum_{j\in J}\left[\left(\frac{1}{r_{(i,j)}}\right)^{12} - \left(\frac{1}{r_{(i,j)}}\right)^{6}\right]$ $r_{(i,j)} = \|x_i - x_j\|$, $x_* \in \mathbb{R}^2$

Ground state configuration given by $\min \sum_{I\cap J=0} E(I,J) \Rightarrow$ 1500 variables to optimize

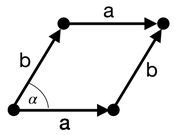
Each frame is an output of a Sequential Quadratic Programming run

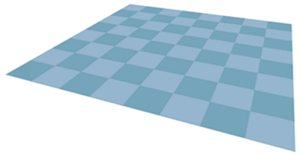
- Random initial configuration
- Boundaries are reduced after each iteration

Ground state configuration is given only by translates of the molecule by a lattice.



For ground state prediction it is enough to optimize 3 variables $[a, b, \alpha]$ instead of 1500.





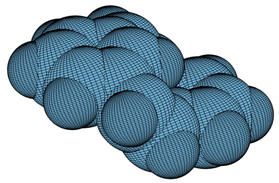
2. Case Study: Molecular Crystal Structure Prediction.

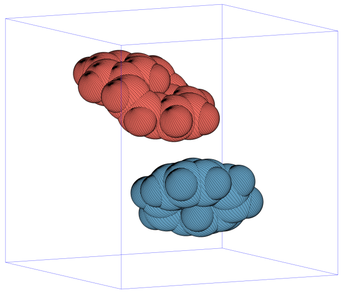
As usual, things are more complicated.

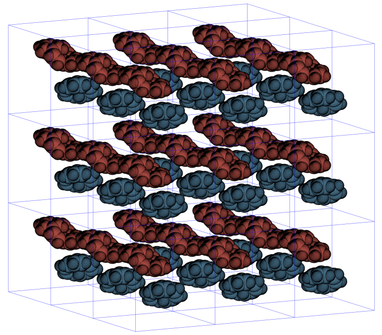
Close - Packing Principle

More symmetry than a simple lattice translation

Molecule as a collection of Van der Waals spheres







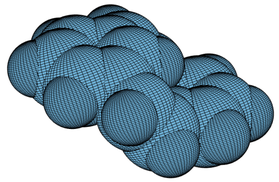
2. Case Study: Molecular Crystal Structure Prediction.

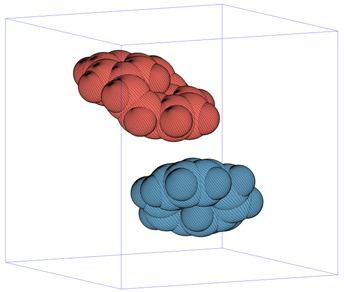
As usual, things are more complicated.

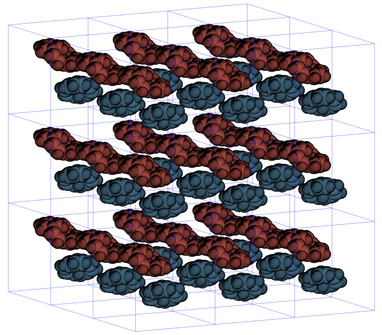
Close - Packing Principle

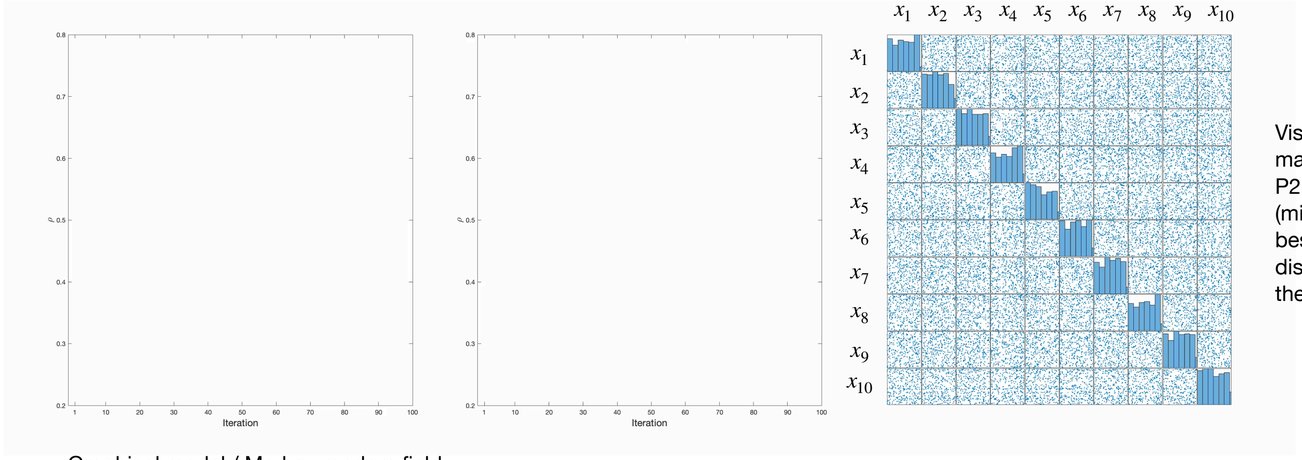
More symmetry than a simple lattice translation

Molecule as a collection of Van der Waals spheres



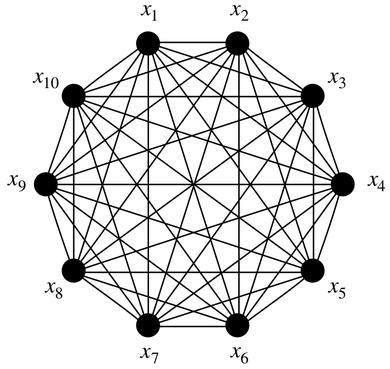






Visualization of the MCYPDE packing maximization run for the space group $P2_1$ (Left) Maximum packing density; (middle) average packing density of N-best packings at $i-th$ iteration; (right) distribution of the packing generation at the iteration.

Graphical model / Markov random field



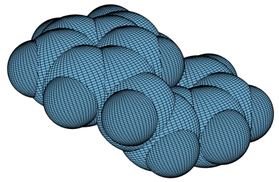
2. Case Study: Molecular Crystal Structure Prediction.

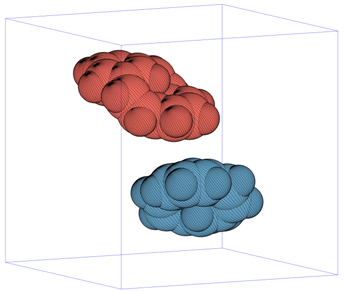
As usual, things are more complicated.

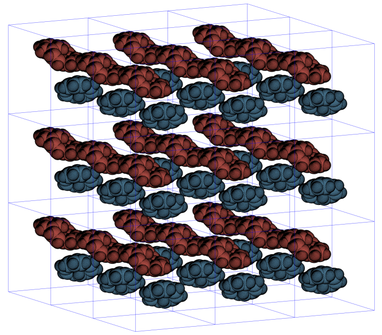
Close - Packing Principle

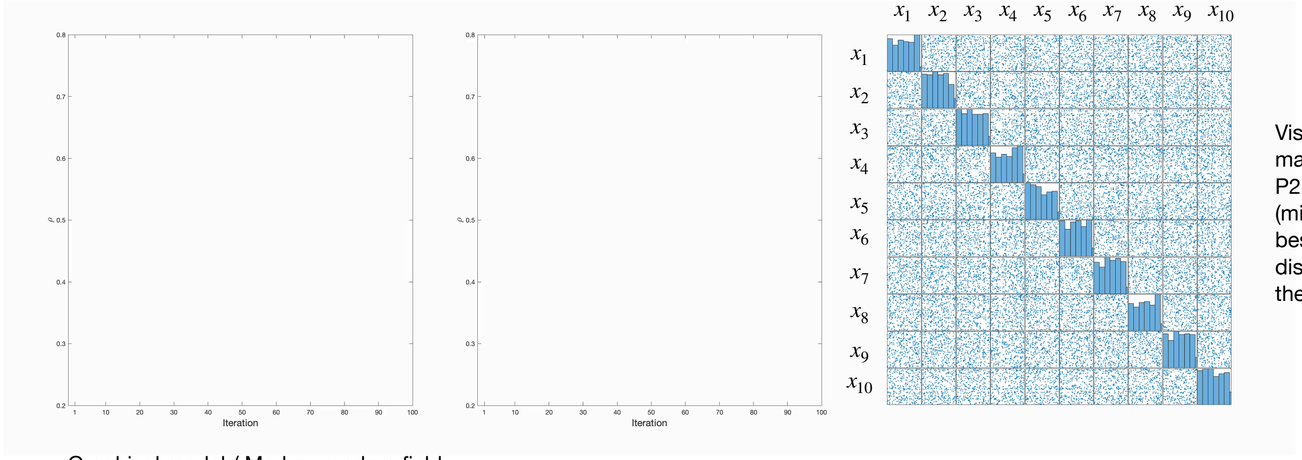
More symmetry than a simple lattice translation

Molecule as a collection of Van der Waals spheres



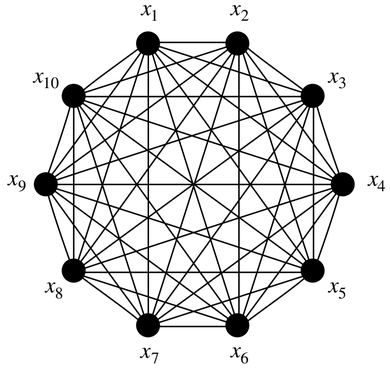






Visualization of the MCYPDE packing maximization run for the space group $P2_1$ (Left) Maximum packing density; (middle) average packing density of N-best packings at $i-th$ iteration; (right) distribution of the packing generation at the iteration.

Graphical model / Markov random field



Belief Propagation / Sum–Product Message Passing

Each node exchanges its belief about the expected value through branches.

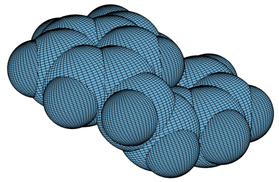
2. Case Study: Molecular Crystal Structure Prediction.

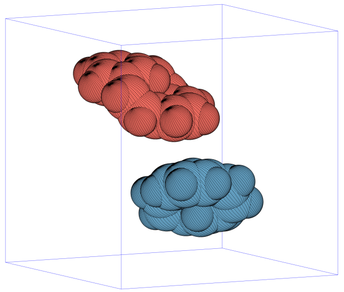
As usual, things are more complicated.

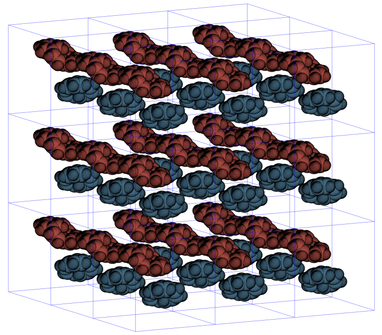
Close - Packing Principle

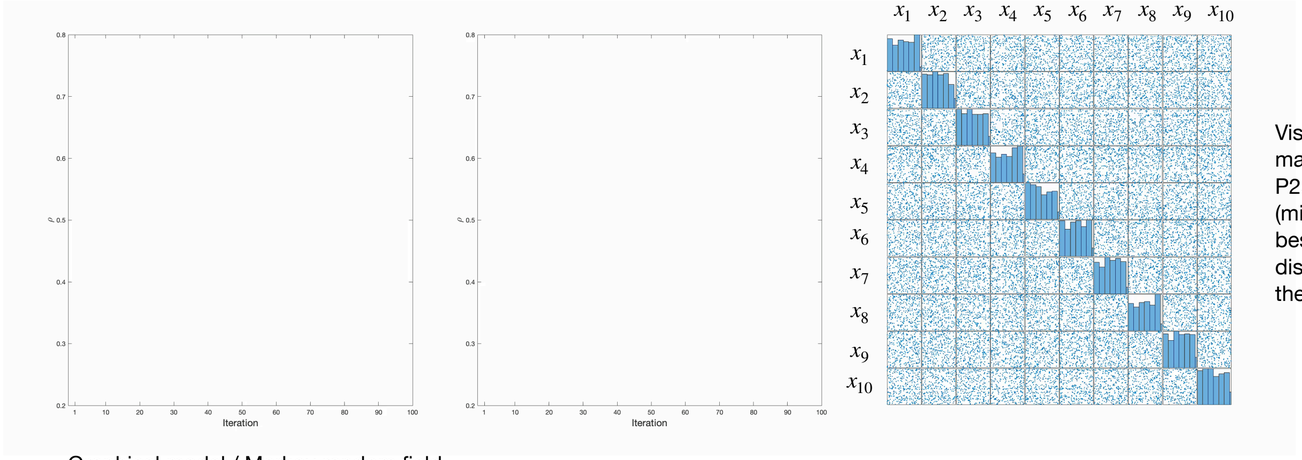
More symmetry than a simple lattice translation

Molecule as a collection of Van der Waals spheres



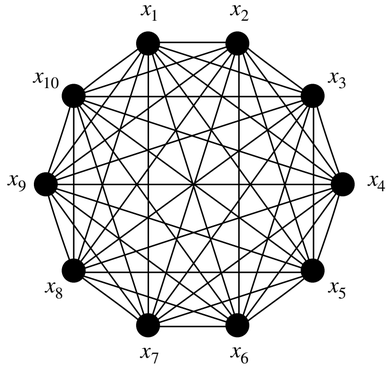






Visualization of the MCYPDE packing maximization run for the space group $P2_1$ (Left) Maximum packing density; (middle) average packing density of N-best packings at $i-th$ iteration; (right) distribution of the packing generation at the iteration.

Graphical model / Markov random field



Belief Propagation / Sum–Product Message Passing

Each node exchanges its belief about the expected value through branches.

Maximization of Multi-Information/Total Correlation

- A measure of stochastic dependence in complex systems.
- Generalization of the **Infomax Principle:**
- A rule for training Artificial Neural Networks.
- The algorithm learns the optimization landscape given by the molecular Close-Packing.

$$\max_{\{f \in \mathcal{F} \mid O = f(I)\}} MI(I;O)$$

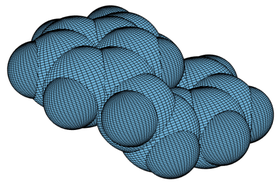
2. Case Study: Molecular Crystal Structure Prediction.

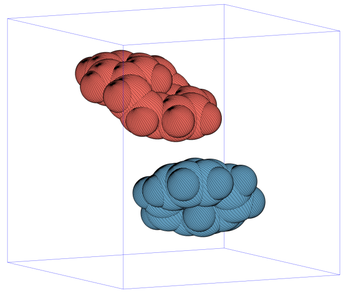
Close - Packing Principle

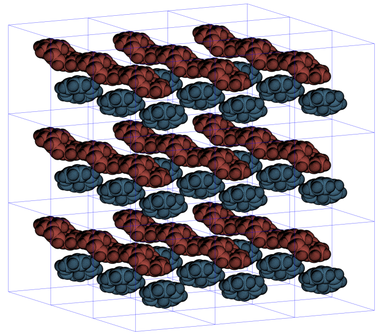
As usual, things are more complicated.

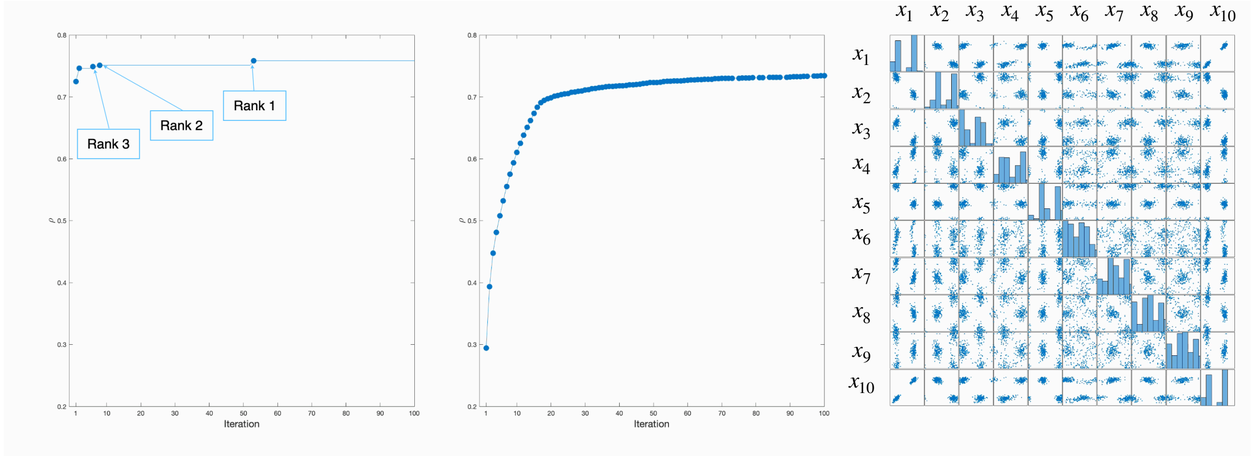
More symmetry than a simple lattice translation

Molecule as a collection of Van der Waals spheres



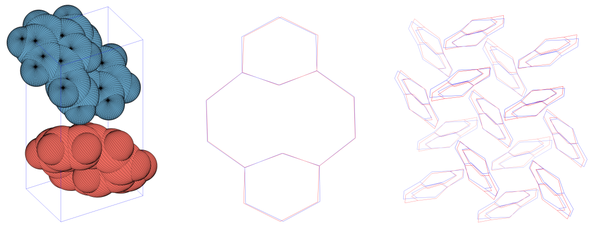






Visualization of the MCYPDE packing maximization run for the space group $P2_1$ (Left) Maximum packing density; (middle) average packing density of N-best packings at $i-th$ iteration; (right) distribution of the packing generation at the iteration.

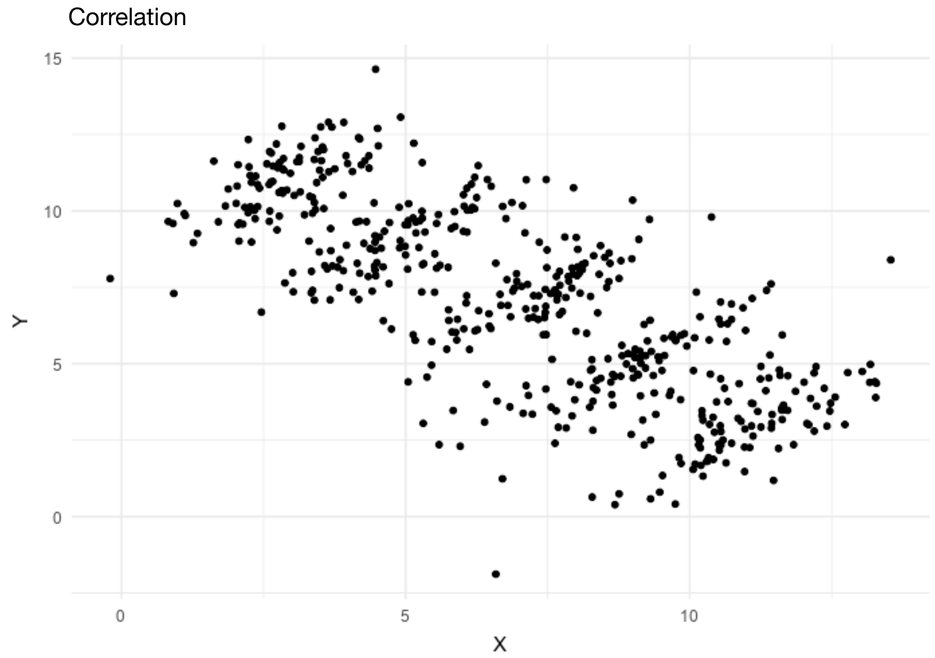
Reference	Comparison	Molecules in Common	RMS	ρ
Ground state	Rank 1	30 out of 30	0.247	0.7584
Ground state	Rank 2	8 out of 30	1.790	0.7514
Ground state	Rank 3	10 out of 30	0.756	0.7510



Visualization of the packing match to the CSP lowest energy structure (Left) A single unit cell displayed in a space-filling representation, where colors indicate symmetry operations modulo lattice translations. (Middle and right) Overlay in a wireframe representation: the global energy minimum (blue) and the matching packing (red). (Middle) Asymmetric unit of the configuration; (right) a 15-molecule cluster.

3. Closing remarks.

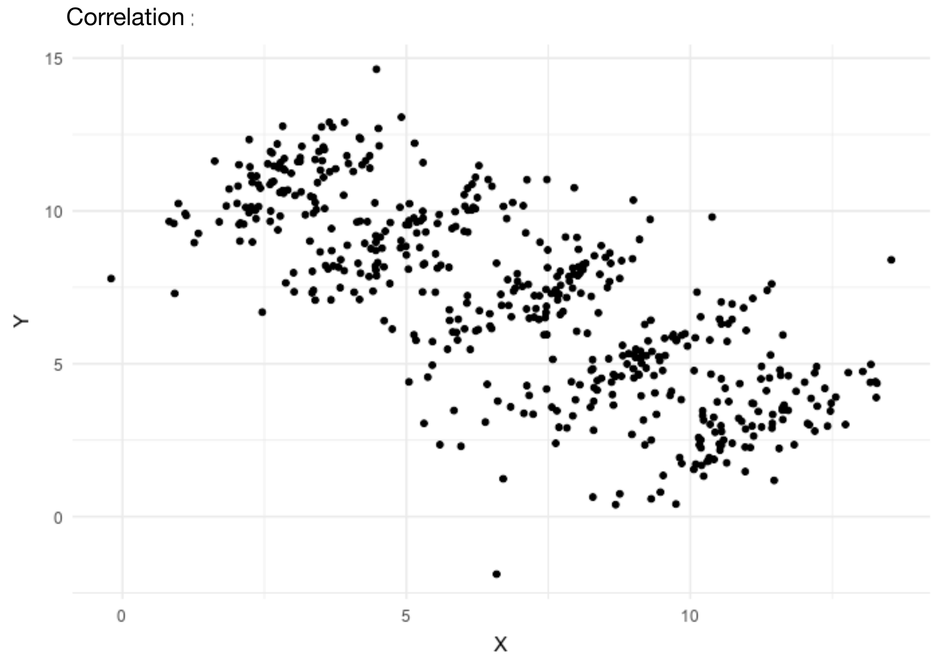
Simpson’s paradox



By Pace~svwiki - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=62007681

3. Closing remarks.

Simpson’s paradox



By Pace~svwiki - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=62007681

Correlation does not imply causation.

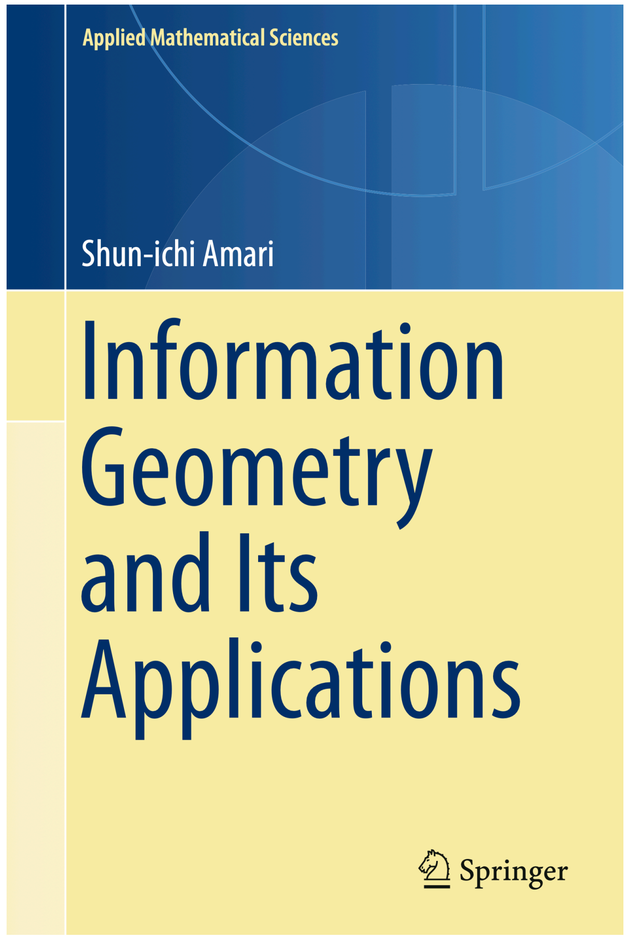
3. Closing remarks.

No free lunch theorem

"state[s] that any two optimization algorithms are equivalent when their performance is averaged across all possible problems”.

Wolpert, D.H.; Macready, W.G. (2005). "Coevolutionary Free Lunches". IEEE Transactions on Evolutionary Computation. 9 (6): 721–735.

3. Closing remarks.



UNIVERSITY OF LIVERPOOL

The Leverhulme Research Centre for Functional Materials Design

LEVERHULME TRUST

THANK YOU

milotorda.net

The Cambridge Crystallographic Data Centre

diamond

Hartree Centre
Science & Technology Facilities Council

ICNS
INTERNATIONAL CENTER FOR NANO-SYSTEMS

Imperial College London

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

UNIVERSITY OF Southampton

The authors thank the Leverhulme Trust for funding this research via the Leverhulme Research Centre for Functional Materials Design