Recovering unknown governing equations from measurement data

Rui Carvalho [1,2,3]

September 29, 2019

[1]Department of Engineering, Durham University, Lower Mountjoy, South Road, Durham, DH1 3LE, UK

[2]Durham Energy Institute, Durham University, South Road, Durham, DH1 3LE, UK

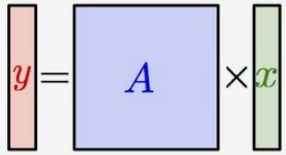
[3]Institute for Data Science, Durham University, South Road, Durham, DH1 3LE, UK

Linear regression in a nutshell

Linear systems in a nutshell

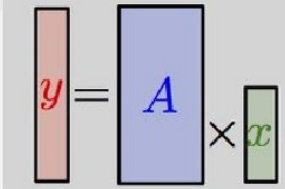
Solving $y \underset{\approx}{=} A\,x \in \mathbb{R}^m \qquad A \in \mathbb{R}^{m \times n}$

Determined $(m = n)$: $x = A^{-1}y$



Over-determined $(m > n)$: $\min_x \|Ax - y\|^2$

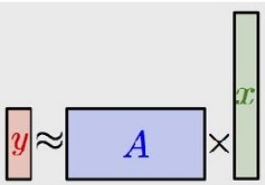
$x = (A^\top A)^{-1} A^\top y \overset{\text{def.}}{=} A^+ y$



least squares

Under-determined $(m < n)$: $\min_x \{\|x\| \;;\; Ax = y\}$

$x = A^\top (AA^\top)^{-1} y \overset{\text{def.}}{=} A^+ y$

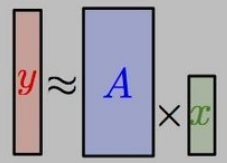


least norm problem

A ill-posed and/or noise: $\min_x \|Ax - y\|^2 + \lambda \|x\|^2$

$x = (A^\top A + \lambda \mathrm{Id}_n)^{-1} A^\top y \overset{\lambda \to 0}{\longrightarrow} A^+ y$

$= A^\top (AA^\top + \lambda \mathrm{Id}_m)^{-1} y$ (Woodbury identity)



modified from Gabriel Peyré

Least squares problem

Least squares problem: choose x to minimise $f(x) = \|Ax - b\|_2^2$ where $A \in \mathbf{R}^{m \times n}$ with $m \geq n$, and $b \in \mathbf{R}^m$ are problem data.

- $m \times n$ matrix A is tall, so $Ax = b$ is over-determined.
- For most choices of b, there is no x that satisfies $Ax = b$.
- *Residual*: $r = Ax - b$.
- Idea: make residual as small as possible, if not 0.
- Assume that the columns of A are independent (the Gram matrix $A^T A$ is invertible), the least-squares approximation problem has the unique solution:

$$x = (A^T A)^{-1} A^T b. \tag{1}$$

- Compare with the solution of the square invertible system $Ax = b$:

$$x = A^{-1} b \tag{2}$$

Regularised approximation

how well the data agrees with the model $Ax=b$

$$\text{minimise} \quad (\overbrace{\|Ax-b\|}, \underbrace{\|x\|}) \tag{3}$$

how large are your model parameters

- $A \in \mathbf{R}^{m \times n}$ is a matrix of n predictors;
- $x \in \mathbf{R}^n$ are the parameters;
- $b \in \mathbf{R}^m$ is a vector of responses.

Idea:

- We want a good fit of $Ax = b$, but we want to do it efficiently, *i.e.*, with small $\|x\|$, so we add to the objective a term that penalises large x.
- Regularisation avoids large x.

The Lasso

Lasso (*least absolute shrinkage and selection operator*)

$$\text{minimise} \quad \|Ax - b\|_2^2 + \lambda \|x\|_1 \tag{4}$$

- Standardise A, so that each column has zero mean and unit variance.
- Solution for $\lambda > 0$ traces out optimal trade-off curve (sweep λ from 0 to ∞).
- Convex problem, so we know how to solve it efficiently.

Can also be written as:

$$\underbrace{\sum_{i=1}^{n} \left(y_i - \overbrace{\beta_0}^{\text{intercept}} - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2}_{\text{Residual Sum of Squares}} + \lambda \sum_{j=1}^{p} |\beta_j| \tag{5}$$

Understanding the energy consumption of electric vehicle charging stations

- Electric Vehicle (EV) charging
  - ELAAD
- Socio-economic, demographic, built environment and land use
  - Population cores.
  - Ambient population grid $1000m \times 1000m$, Landscan 2012.
  - Corine Land Use and Land Cover.
  - Neighbourhoods data.
  - Energy Atlas
  - Traffic flows
  - OSM amenities

ELAAD raw dataset

- ElaadNL: Dutch research organisation involved in the development and deployment of EV charging technologies.
- $1,747$ georeferenced charging stations.
- $54,000$ users, each identified by a unique *id*.
- $1,060,763$ charging events.
- Data collected between January 2012 and March 2016.

ELAAD: spatial data

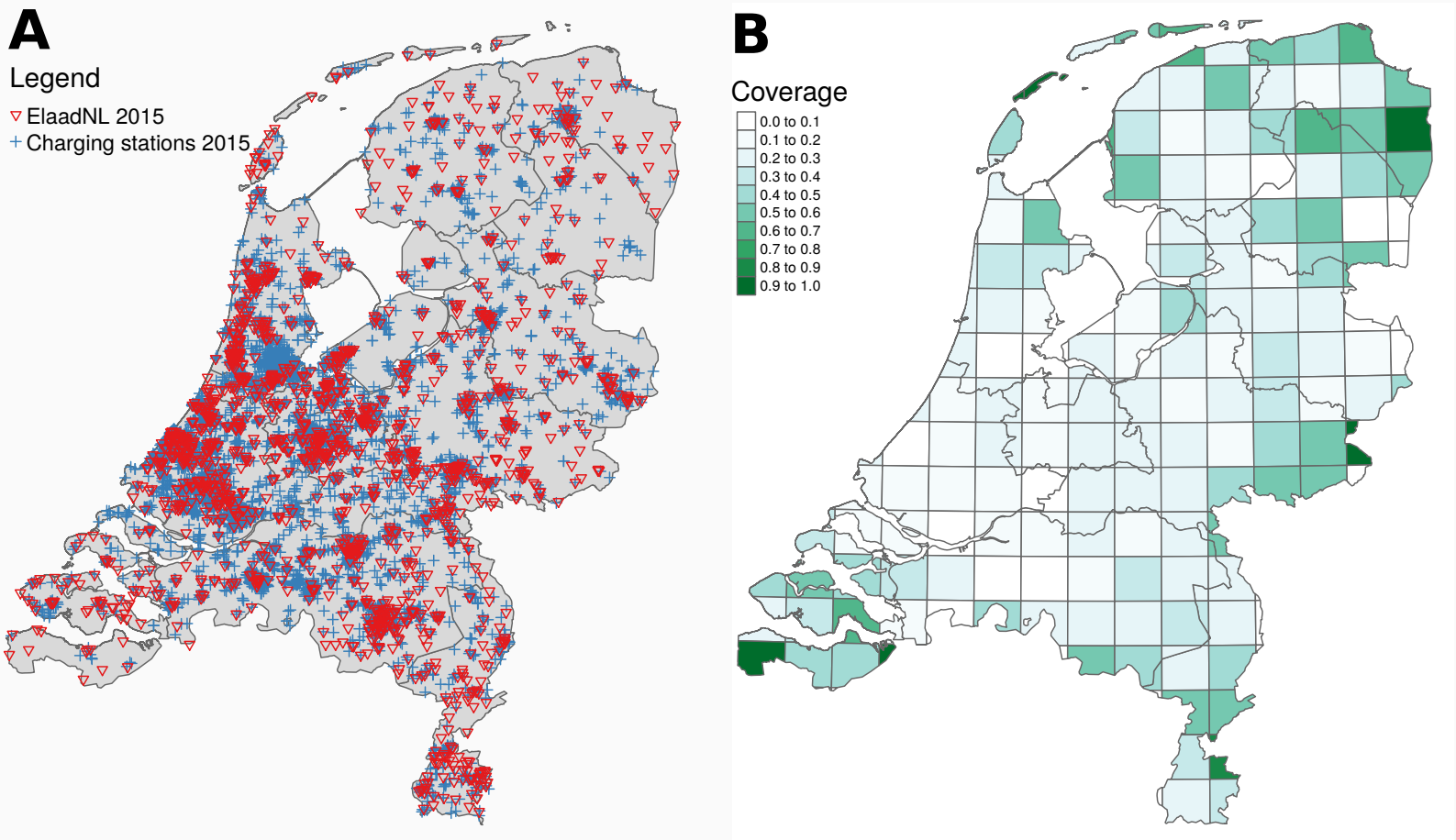


Figure 1: Public charging stations in the ElaadNL data set in 2015 (triangles) shown together with the Charging stations 2015 dataset (crosses). In the year 2015, 17 786 publicly available connectors for slow charging were operational in the NL. We identified 8 400 unique positions of charging stations, i.e. considering the distribution of connectors at charging stations observed in the ElaadL dataset, this data covers 78.3% of all stations. In panel (B), we estimated the spatial representativeness of the ElaadNL data sets by calculating the ratio between the number of station in ELaadNL and in Charging stations 2015 located in squared cells of a regular grid. In the largest cities, Amsterdam and Rotterdam, the data contains a small percentage of all charging stations.

Predictors: GIS data

- **Vector data**
  - **Polygon data**
    - population cores,
    - neighbourhoods data,
    - energy atlas,
    - liveability,
    - land use and land cover (urban atlas, CBS land cover).
  - **Polyline data**
    - traffic flow data.
  - **Point data**
    - OSM amenities,
    - OpenChargeMap.
- **Raster data**
  - LandScan - ambient population.

Vector polygon data: Population cores (2168 cores)

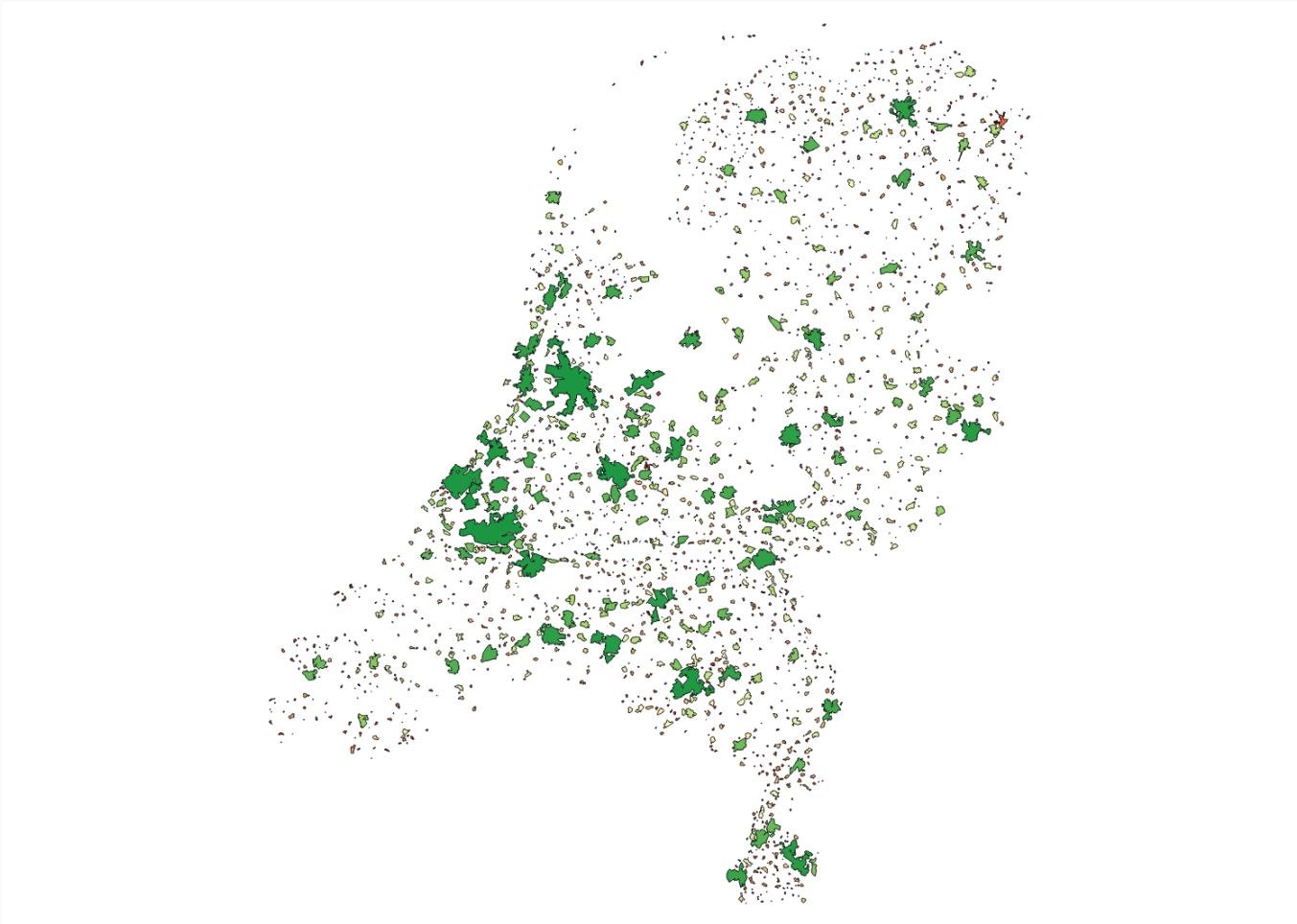


Figure 2: Population cores are continuous spatial units with at least 25 homes or 50 inhabitants (102 predictors). Source: Statistics Netherlands https://opendata.cbs.nl/

Polygon data: Population cores (2168 cores, 84 attributes)

- number of persons in private household,
- number of persons in private households, 0 to 15 years,
- number of persons in private households, 15 to 25 years,
- number of persons in private households 25 to 45 years,
- number of persons in private households, 45 to 65 years,
- number of persons in private households 65 years or older,
- number of persons in one-person households,
- number of people in multi-person household with children,
- number of people in multi-person household without children,
- nercentage of working population, 15-24 years old,
- number of households of two persons,
- number of households of three persons,
- the number of residential units,
- ...

Lasso

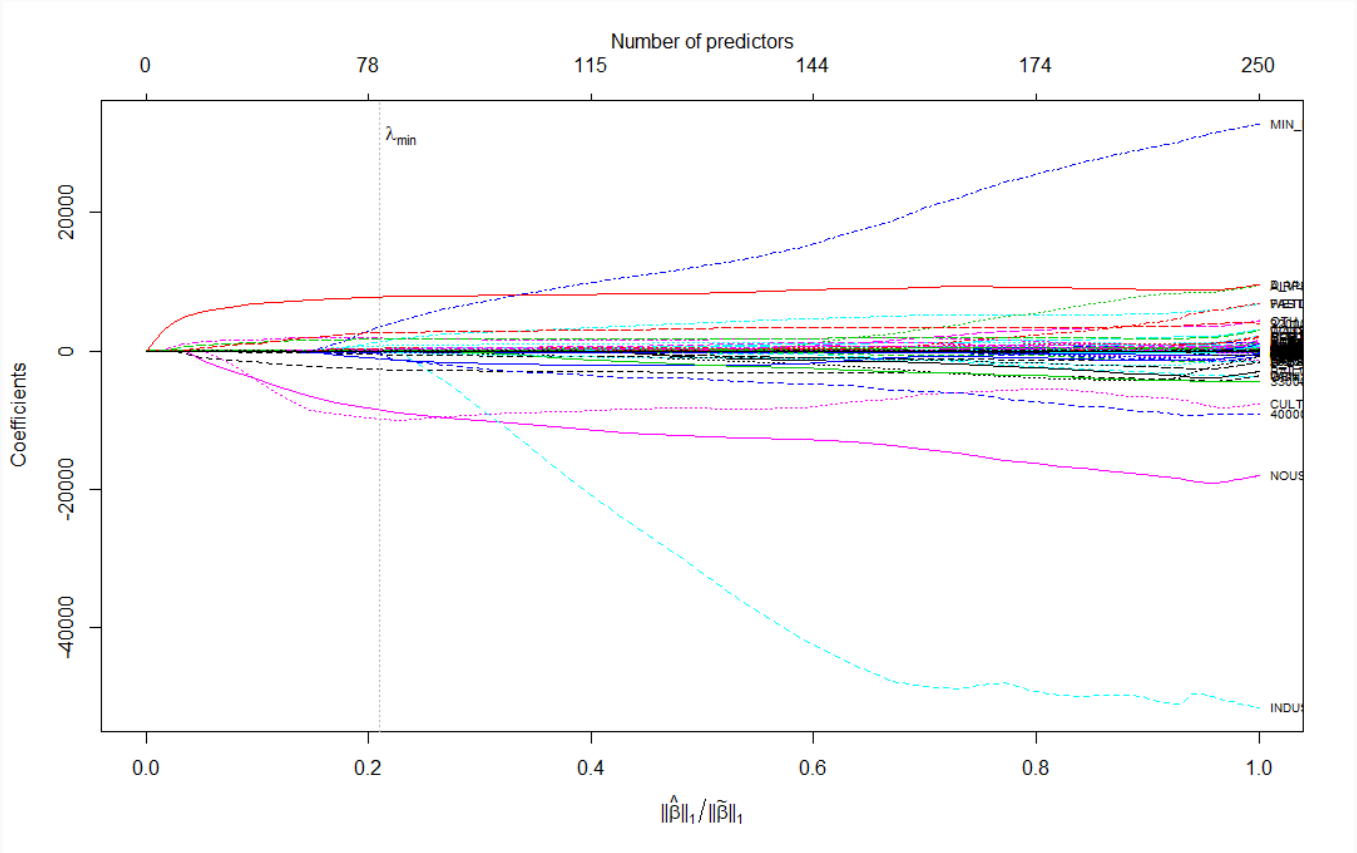
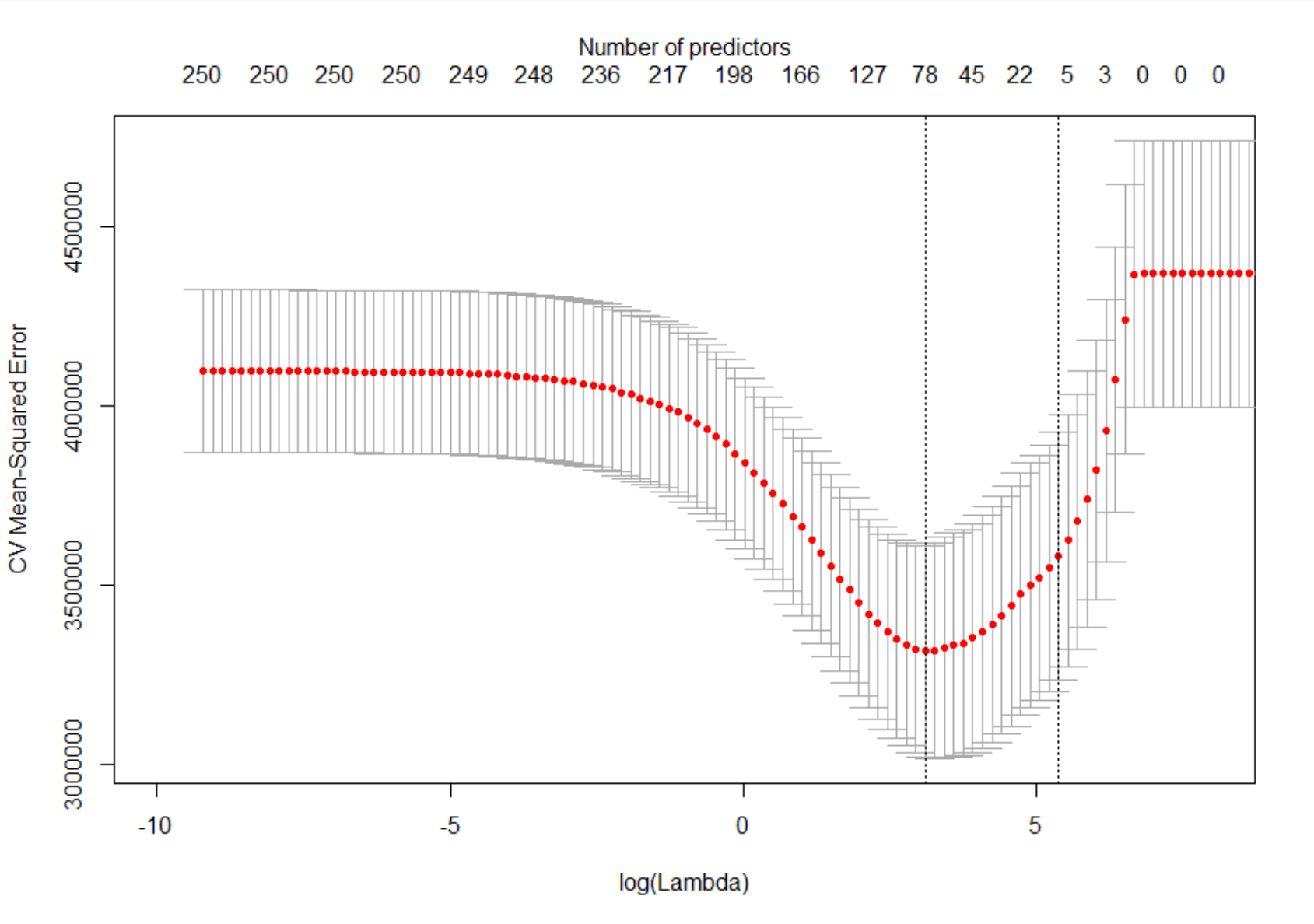


Figure 3: Lasso

$$\underbrace{\sum_{i=1}^{n}\left(y_i - \overbrace{\beta_0}^{\text{intercept}} - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2}_{\text{Residual Sum of Squares}} + \lambda \sum_{j=1}^{p}|\beta_j| \tag{6}$$

Lasso


Number of predictors
250 250 250 250 249 248 236 217 198 166 127 78 45 22 5 3 0 0 0
CV Mean-Squared Error
4500000
4000000
3500000
3000000
-10
-5
0
5
log(Lambda)

Lasso fit

- Lasso does variable selection on 240 predictors.
- At the optimal λ, we reduce the number of predictors to 79 (about 1/3 of the original predictors).
- $R^2 = 0.362$ at optimum value ($R^2(adjusted) = 0.316$).
- The λ_{min} is the one which minimises the error in cross-validation. The $\lambda_{1\sigma}$ is the λ value within 1 standard error of λ_{min}.

Coefficient	Meaning
-27.8627	The percentage of working population working in Mining, Manufacturing and Construction.
20.3313	The percentage of working population employed in commercial services
18.169	The percentage working population engaged in agriculture, forestry and fisheries, industry, commercial and non-commercial services
-0.1884	The percentage of the number of multi-person households without children
4.1384	Number of business Services
14.5717	Property unknown (no link between the addresses of the Key Registers Addresses and the housing register Cadaster).
57.8468	Average income per inhabitant
-0.9093	Average distance of all residents in an area to the nearest shops for groceries.

Data-driven discovery of dynamical systems

Data-driven discovery of dynamical systems

Goal of computationally-oriented scientists:

> Inferring a (typically nonlinear) model from observations that both correctly identifies the underlying dynamics *and* generalises qualitatively and quantitatively to unmeasured parts of the phase, parameter, or application space.

- ODE or PDE system described by

$$u_t = N(u, x, t; \overbrace{\mu}^{\text{parameter}}) \tag{7}$$

- Our objective is to discover $N(.)$ given only time-series measurements of the system.
- A key assumption (prior) is that the true $N(.)$ is comprised of only a few terms, making the model sparse in the space of all possible combinations of functions.
- For example, Burgers' equation

$$N = -uu_x + \mu u_{xx} \tag{8}$$

and the harmonic oscillator

$$N = -i\mu x^2 u - i\hbar u_{xx}/2 \tag{9}$$

each have only two terms.

PNAS

Discovering governing equations from data by sparse identification of nonlinear dynamical systems

Steven L. Brunton[a,1], Joshua L. Proctor[b], and J. Nathan Kutz[c]

[a]Department of Mechanical Engineering, University of Washington, Seattle, WA 98195; [b]Institute for Disease Modeling, Bellevue, WA 98005; and [c]Department of Applied Mathematics, University of Washington, Seattle, WA 98195

Edited by William Bialek, Princeton University, Princeton, NJ, and approved March 1, 2016 (received for review August 31, 2015)

Extracting governing equations from data is a central challenge in many diverse areas of science and engineering. Data are abundant whereas models often remain elusive, as in climate science, neuroscience, ecology, finance, and epidemiology, to name only a few examples. In this work, we combine sparsity-promoting techniques and machine learning with nonlinear dynamical systems to discover governing equations from noisy measurement data. The only assumption about the structure of the model is that there are only a few important terms that govern the dynamics, so that the equations are sparse in the space of possible functions; this assumption holds for many physical systems in an appropriate basis. In particular, we use sparse regression to determine the fewest terms in the dynamic governing equations required to accurately represent the data. This results in parsimonious models that balance accuracy with model complexity to avoid overfitting. We demonstrate the algorithm on a wide range of problems, from simple canonical systems, including linear and nonlinear oscillators and the chaotic Lorenz system, to the fluid vortex shedding behind an obstacle. The fluid example illustrates the ability of this method to discover the underlying dynamics of a system that took experts in the community nearly 30 years to resolve. We also show that this method generalizes to parameterized systems and systems that are time-varying or have external forcing.

dynamical systems | machine learning | sparse regression | system identification | optimization

dynamical systems from data. However, symbolic regression is expensive, does not scale well to large systems of interest, and may be prone to overfitting unless care is taken to explicitly balance model complexity with predictive power. In ref. 4, the Pareto front is used to find parsimonious models. There are other techniques that address various aspects of the dynamical system discovery problem. These include methods to discover governing equations from time-series data (6), equation-free modeling (7), empirical dynamic modeling (8, 9), modeling emergent behavior (10), and automated inference of dynamics (11–13); ref. 12 provides an excellent review.

Sparse Identification of Nonlinear Dynamics (SINDy)

In this work, we reenvision the dynamical system discovery problem from the perspective of sparse regression (14–16) and compressed sensing (17–22). In particular, we leverage the fact that most physical systems have only a few relevant terms that define the dynamics, making the governing equations sparse in a high-dimensional nonlinear function space. The combination of sparsity methods in dynamical systems is quite recent (23–30). Here, we consider dynamical systems (31) of the form

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)). \qquad [1]$$

The vector $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the state of a system at time t, and the function $\mathbf{f}(\mathbf{x}(t))$ represents the dynamic constraints that de-

Data-driven discovery of dynamical systems: method

Method:

- Construct a library $\Theta(U)$ of candidate linear, nonlinear, and partial derivative terms for the right-hand side.
- Each column of $\Theta(U)$ contains the values of a candidate term evaluated using the collected data.
- In this library, one can write the dynamics as

$$U_t = \Theta(U)\xi \tag{10}$$

  where
  - U_t is a vector of time derivatives of the measurement data.
  - ξ is a sparse vector, with each nonzero entry corresponding to a functional term to be included in the dynamics.
- Finding the sparsest vector ξ consistent with the measurement data is now feasible with advanced methods in sparse regression, which makes it possible to find the most parsimonious model while circumventing a combinatorial search.

Identifying the Lorenz Equations

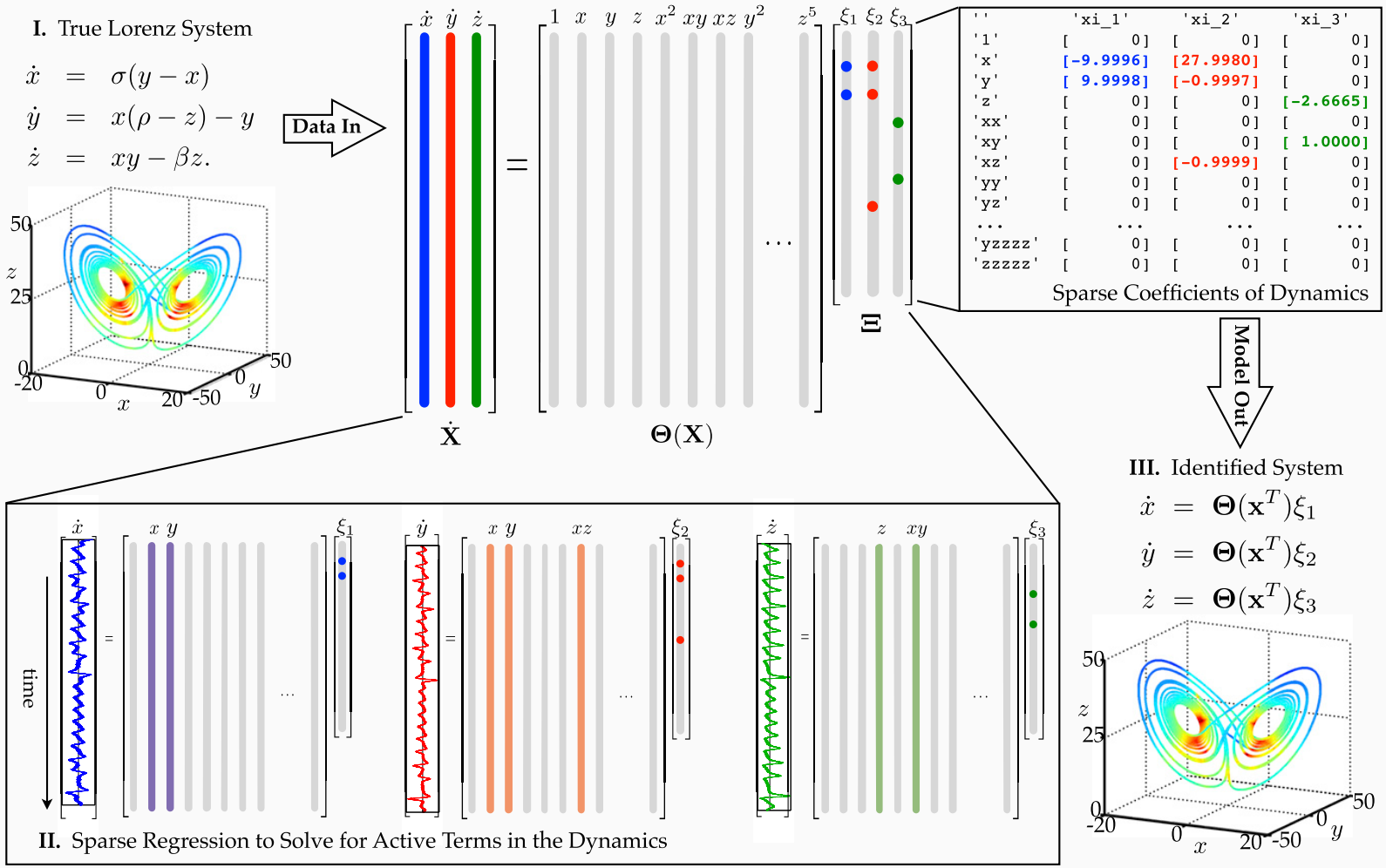


Fig. 1. Schematic of the SINDy algorithm, demonstrated on the Lorenz equations. Data are collected from the system, including a time history of the states $\mathbf{X}$ and derivatives $\dot{\mathbf{X}}$; the assumption of having $\dot{\mathbf{X}}$ is relaxed later. Next, a library of nonlinear functions of the states, $\Theta(\mathbf{X})$, is constructed. This nonlinear feature library is used to find the fewest terms needed to satisfy $\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi$. The few entries in the vectors of Ξ, solved for by sparse regression, denote the relevant terms in the right-hand side of the dynamics. Parameter values are $\sigma = 10, \beta = 8/3, \rho = 28$, $(x_0, y_0, z_0)^T = (-8, 7, 27)^T$. The trajectory on the Lorenz attractor is colored by the adaptive time step required, with red indicating a smaller time step.