# Glossary

**Adjoint –** For a finite-dimensional linear map (i.e., a matrix $\mathbf{A}$), the adjoint $\mathbf{A}^*$ is given by the complex conjugate transpose of the matrix. In the infinite-dimensional context, the adjoint $\mathcal{A}^*$ of a linear operator $\mathcal{A}$ is defined so that $\langle \mathcal{A} f, g \rangle = \langle f, \mathcal{A}^* g \rangle$, where $\langle \cdot, \cdot \rangle$ is an inner product.

**Akaike information criterion (AIC) –** An estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.

**Backpropagation (Backprop) –** A method used for computing the gradient descent required for the training of neural networks. Based upon the chain rule, backprop exploits the compositional nature of NNs in order to frame an optimization problem for updating the weights of the network. It is commonly used to train deep neural networks.

**Balanced input–output model –** A model expressed in a coordinate system where the states are ordered hierarchically in terms of their joint controllability and observability. The controllability and observability Gramians are equal and diagonal for such a system.

**Bayesian information criterion (BIC) –** An estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, BIC estimates the quality of each model, relative to each of the other models. Thus, BIC provides a means for model selection.

**Classification –** A general process related to categorization, the process in which ideas and objects are recognized, differentiated, and understood. Classification is a common task for machine learning algorithms.

**Closed-loop control –** A control architecture where the actuation is informed by sensor data about the output of the system.

**Clustering –** A task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a primary goal of exploratory data mining, and a common technique for statistical data analysis.

**Coherent structure –** A spatial mode that is correlated with the data from a system.

**Compression –** The process of reducing the size of a high-dimensional vector or array by approximating it as a sparse vector in a transformed basis. For example, MP3 and JPG compression use the Fourier basis or Wavelet basis to compress audio or image signals.

**Compressed sensing –** The process of reconstructing a high-dimensional vector signal from a random under sampling of the data using the fact that the high-dimensional signal is sparse in a known transform basis, such as the Fourier basis.

**Control theory –** The framework for modifying a dynamical system to conform to desired engineering specification through sensing and actuation.

**Controllability –** A system is controllable if it is possible to steer the system to any state with actuation. Degrees of controllability are determined by the controllability Gramian.

**Convex optimization –** An algorithmic frameworks for minimizing convex functions over convex sets.

**Convolutional neural network (CNN) –** A class of deep, feed-forward neural networks that is especially amenable to analyzing natural images. The convolution is typically a spatial filter which synthesizes local (neighboring) spatial information.

**Cross-validation –** A model validation technique for assessing how the results of a statistical analysis will generalize to an independent (withheld) data set.

**Data matrix –** A matrix where each column vector is a snapshot of the state of a system at a particular instance in time. These snapshots may be *sequential* in time, or they may come from an ensemble of initial conditions or experiments.

**Deep learning –** A class of machine learning algorithms that typically uses deep CNNs for feature extraction and transformation. Deep learning can leverage supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) algorithms, learning multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

**DMD amplitude –** The amplitude of a given DMD mode as expressed in the data. These amplitudes may be interpreted as the significance of a given DMD mode, similar to the power spectrum in the FFT.

**DMD eigenvalue –** Eigenvalues of the best-fit DMD operator $\mathbf{A}$ (see *dynamic mode decomposition*) representing an oscillation frequency and a growth or decay term.

**DMD mode (also *dynamic mode*) –** An eigenvector of the best-fit DMD operator $\mathbf{A}$ (see *dynamic mode decomposition*). These modes are spatially coherent and oscillate in time at a fixed frequency and a growth or decay rate.

**Dynamic mode decomposition (DMD) –** The leading eigendecomposition of a best-fit linear operator $\mathbf{A} = \mathbf{X}'\mathbf{X}^{\dagger}$ that propagates the data matrix $\mathbf{X}$ into a future data matrix $\mathbf{X}'$. The eigenvectors of $\mathbf{A}$ are DMD modes and the corresponding eigenvalues determine the time dynamics of these modes.

**Dynamical system –** A mathematical model for the dynamic evolution of a system. Typically, a dynamical system is formulated in terms of ordinary differential equations on a state-space. The resulting equations may be linear or nonlinear and may also include the effect of actuation inputs and represent outputs as sensor measurements of the state.

**Eigensystem realization algorithm (ERA) –** A system identification technique that produces balanced input–output models of a system from impulse response data. ERA has been shown to produce equivalent models to balanced proper orthogonal decomposition and dynamic mode decomposition under some circumstances.

**Emission –** The measurement functions for a hidden Markov model.

**Feedback control –** Closed-loop control where sensors measure the downstream effect of actuators, so that information is fed back to the actuators. Feedback is essential for robust control where model uncertainty and instability may be counteracted with fast sensor feedback.

**Feedforward control –** Control where sensors measure the upstream disturbances to a system, so that information is fed forward to actuators to cancel disturbances proactively.

**Fast Fourier transform (FFT) –** A numerical algorithm to compute the discrete Fourier transform (DFT) in $\mathcal{O}(n \log(n))$ operations. The FFT has revolutionized modern computations, signal processing, compression, and data transmission.

**Fourier transform –** A change of basis used to represent a function in terms of an infinite series of sines and cosines.

**Galerkin projection –** A process by which governing partial differential equations are reduced into ordinary differential equations in terms of the dynamics of the coefficients of a set of orthogonal basis modes that are used to approximate the solution.

**Gramian –** The controllability (resp. observability) Gramian determines the degree to which a state is controllable (resp. observable) via actuation (resp. via estimation). The Gramian establishes an inner product on the state space.

**Hidden Markov model (HMM) –** A Markov model where there is a hidden state that is only observed through a set of measurements known as emissions.

**Hilbert space –** A generalized vector space with an inner product. When referred to in this text, a Hilbert space typically refers to an infinite-dimensional function space. These spaces are also complete metric spaces, providing a sufficient mathematical framework to enable calculus on functions.

**Incoherent measurements –** Measurements that have a small inner product with the basis vectors of a sparsifying transform. For instance, single pixel measurements (i.e., spatial delta functions) are incoherent with respect to the spatial Fourier transform basis, since these single pixel measurements excite all frequencies and do not preferentially align with any single frequency.

**Kalman filter –** An estimator that reconstructs the full state of a dynamical system from measurements of a time-series of the sensor outputs and actuation inputs. A Kalman filter is itself a dynamical system that is constructed for observable systems to stably converge to the true state of the system. The Kalman filter is optimal for linear systems with Gaussian process and measurement noise of a known magnitude.

**Koopman eigenfunction –** An eigenfunction of the Koopman operator. These eigenfunctions correspond to measurements on the state-space of a dynamical system that form intrinsic coordinates. In other words, these intrinsic measurements will evolve linearly in time despite the underlying system being nonlinear.

**Koopman operator –** An infinite-dimensional linear operator that propagates measurement functions from an infinite dimensional Hilbert space through a dynamical system.

**Least squares regression –** A regression technique where a best-fit line or vector is found by minimizing the sum of squares of the error between the model and the data.

**Linear quadratic regulator (LQR) –** An optimal proportional feedback controller for full-state feedback, which balances the objectives of regulating the state while not expending too much control energy. The proportional gain matrix is determined by solving an algebraic Riccati equation.

**Linear system –** A system where superposition of any two inputs results in the superposition of the two corresponding outputs. In other words, doubling the input doubles the output. Linear time-invariant dynamical systems are characterized by linear operators, which are represented as matrices.

**Low rank –** A property of a matrix where the number of linearly independent rows and columns is small compared with the size of the matrix. Generally, low-rank approximations are sought for large data matrices.

**Machine learning –** A set of statistical tools and algorithms that are capable of extracting the dominant patterns in data. The data mining can be supervised or unsupervised, with the goal of clustering, classification and prediction.

**Markov model –** A probabilistic dynamical system where the state vector contains the probability that the system will be in a given state; thus, this state vector must always sum to unity. The dynamics are given by the Markov transition matrix, which is constructed so that each row sums to unity.

**Markov parameters –** The output measurements of a dynamical system in response to an impulsive input.

**Max pooling –** A data down-sampling strategy whereby an input representation (image, hidden-layer output matrix, etc.) is reduced in dimensionality, thus allowing for assumptions to be made about features contained in the down-sampled sub-regions.

**Model predictive control (MPC) –** A form of optimal control that optimizes a control policy over a finite-time horizon, based on a model. The models used for MPC are typically linear and may be determined empirically via system identification.

**Moore's law –** The observation that transistor density, and hence processor speed, increases exponentially in time. Moore's law is commonly used to predict future computational power and the associated increase in the scale of problem that will be computationally feasible.

**Multiscale –** The property of having many scales in space and/or time. Many systems, such as turbulence, exhibit spatial and temporal scales that vary across many orders of magnitude.

**Observability –** A system is observable if it is possible to estimate any system state with a time-history of the available sensors. Degrees of observability are determined by the observability Gramian.

**Observable function –** A function that measures some property of the state of a system. Observable functions are typically elements of a Hilbert space.

**Optimization –** Generally a set of algorithms that find the "best available" values of some objective function given a defined domain (or input), including a variety of different types of objective functions and different types of domains. Mathematically, optimization aims to maximize or minimize real function by systematically choosing input values from within an allowed set and computing the value of the function. The generalization of optimization theory and techniques to other formulations constitutes a large area of applied mathematics.

**Overdetermined system –** A system $\mathbf{A}\mathbf{x} = \mathbf{b}$ where there are more equations than unknowns. Usually there is no exact solution $\mathbf{x}$ to an overdetermined system, unless the vector $\mathbf{b}$ is in the column space of $\mathbf{A}$.

**Pareto front –** The allocation of resources from which it is impossible to reallocate so as to make any one individual or preference criterion better off without making at least one individual or preference criterion worse off.

**Perron-Frobenius operator –** The adjoint of the Koopman operator, the Perron-Frobenius operator is an infinite-dimensional operator that advances probability density functions through a dynamical system.

**Power spectrum –** The squared magnitude of each coefficient of a Fourier transform of a signal. The power corresponds to the amount of each frequency required to reconstruct a given signal.

**Principal component –** A spatially correlated mode in a given data set, often computed using the singular value decomposition of the data after the mean has been subtracted.

**Principal components analysis (PCA) –** A decomposition of a data matrix into a hierarchy of principal component vectors that are ordered from most correlated to least correlated with the data. PCA is computed by taking the singular value decomposition of the data after subtracting the mean. In this case, each singular value represents the variance of the corresponding principal component (singular vector) in the data.

**Proper orthogonal decomposition (POD) –** The decomposition of data from a dynamical system into a hierarchical set of orthogonal modes, often using the singular value decomposition. When the data consists of velocity measurements of a system, such as an incompressible fluid, then the proper orthogonal decomposition orders modes in terms of the amount of energy these modes contain in the given data.

**Pseudo-inverse –** The pseudo-inverse generalizes the matrix inverse for non-square matrices, and is often used to compute the least-squares solution to a system of equations. The SVD is a common method to compute the pseudo-inverse: given the SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, the pseudo-inverse is $\mathbf{X}^\dagger = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^*$.

**Recurrent neural network (RNN) –** A class of neural networks where connections between units form a directed graph along a sequence. This allows it to exhibit dynamic temporal behavior for a time sequence.

**Reduced-order model (ROM) –** A model of a high-dimensional system in terms of a low-dimensional state. Typically, a reduced-order model balances accuracy with computational cost of the model.

**Regression –** A statistical model that represents an outcome variable in terms of indicator variables. Least-squares regression is a linear regression that finds the line of best fit to data; when generalized to higher dimensions and multi-linear regression, this generalizes to principal components regression. Nonlinear regression, dynamic regression, and functional or semantic regression are used in system identification, model reduction, and machine learning.

**Restricted isometry property (RIP) –** The property that a matrix acts like a unitary matrix, or an isometry map, on sparse vectors. In other words, the distance between any two sparse vectors is preserved if these vectors are mapped through a matrix that satisfies the restricted isometry property.

**Robust control –** A field of control that penalizes *worst case scenario* control outcomes, thus promoting controllers that are robust to uncertainties, disturbances, and unmodeled dynamics.

**Robust statistics –** Methods for producing good statistical estimates for data drawn from a wide range of probability distributions, especially for distributions that are not normal and where outliers compromise predictive capabilities.

**Singular value decomposition (SVD) –** Given a matrix $\mathbf{X} \in \mathbb{C}^{n\times m}$, the SVD is given by $\mathbf{X} = \mathbf{U\Sigma V}^*$ where $\mathbf{U} \in \mathbb{C}^{n\times n}$, $\mathbf{\Sigma} \in \mathbb{C}^{n\times m}$, and $\mathbf{V} \in \mathbb{C}^{m\times m}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ are unitary, so that $\mathbf{UU}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}$ and $\mathbf{VV}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I}$. The matrix $\mathbf{\Sigma}$ has entries along the diagonal corresponding to the singular values that are ordered from largest to smallest. This produces a hierarchical matrix decomposition that splits a matrix into a sum of rank-1 matrices given by the outer product of a column vector (left singular vector) with a row vector (conjugate transpose of right singular vector). These rank-1 matrices are ordered by the singular value so that the first $r$ rank-1 matrices form the *best* rank-$r$ matrix approximation of the original matrix in a least-squares sense.

**Snapshot –** A single high-dimensional measurement of a system at a particular time. A number of snapshots collected at a sequence of times may be arranged as column vectors in a data matrix.

**Sparse identification of nonlinear dynamics (SINDy) –** A nonlinear system identification framework used to simultaneously identify the nonlinear structure and parameters of a dynamical system from data. Various sparse optimization techniques may be used to determine SINDy models.

**Sparsity –** A vector is *sparse* if most of its entries are zero or nearly zero. Sparsity refers to the observation that most data are sparse when represented as vectors in an appropriate transformed basis, such as Fourier or POD bases.

**Spectrogram –** A short-time Fourier transform computed on a moving window, which results in a time-frequency plot of which frequencies are active at a given time. The spectrogram is useful for characterizing nonperiodic signals, where the frequency content evolves over time, as in music.

**State space –** The set of all possible system states. Often the state-space is a vector space, such as $\mathbb{R}^n$, although it may also be a smooth manifold $\mathcal{M}$.

**Stochastic gradient descent –** Also known as incremental gradient descent, it allows one to approximate the gradient with a single data point instead of all available data. At each step of the gradient descent, a randomly chosen data point is used to compute the gradient direction.

**System identification –** The process by which a model is constructed for a system from measurement data, possibly after perturbing the system.

**Time delay coordinates –** An augmented set of coordinates constructed by considering a measurement at the current time along with a number of times in the past at fixed intervals from the current time. Time delay coordinates are often useful in reconstructing attractor dynamics for systems that do not have enough measurements, as in the Takens embedding theorem.

**Total least squares –** A least-squares regression algorithm that minimizes the error on both the inputs and the outputs. Geometrically, this corresponds to finding the line that minimizes the sum of squares of the total distance to all points, rather than the sum of squares of the vertical distance to all points.

**Uncertainty quantification (UQ) –** The principled characterization and management of uncertainty in engineering systems. Uncertainty quantification often involves the application of powerful tools from probability and statistics to dynamical systems.

**Underdetermined system –** A system $\mathbf{A}\mathbf{x} = \mathbf{b}$ where there are fewer equations than unknowns. Generally the system has infinitely many solutions $\mathbf{x}$ unless $\mathbf{b}$ is not in the column space of $\mathbf{A}$.

**Unitary matrix –** A matrix whose complex conjugate transpose is also its inverse. All eigenvalues of a unitary matrix are on the complex unit circle, and the action of a unitary matrix may be through of as a change of coordinates that preserves the Euclidean distance between any two vectors.

**Wavelet –** A generalized function, or family of functions, used to generalize the Fourier transform to approximate more complex and multiscale signals.