Common Optimization Techniques, Equations, Symbols, and Acronyms

Most Common Optimization Strategies

Least-Squares (discussed in Chapters 1 and 4) minimizes the sum of the squares of the residuals between a given fitting model and data. Linear least-squares, where the residuals are linear in the unknowns, has a closed form solution which can be computed by taking the derivative of the residual with respect to each unknown and setting it to zero. It is commonly used in the engineering and applied sciences for fitting polynomial functions. Nonlinear least-squares typically requires iterative refinement based upon approximating the nonlinear least-squares with a linear least-squares at each iteration.

Gradient Descent (discussed in Chapters 4 and 6) is the industry leading, convex optimization method for high-dimensional systems. It minimizes residuals by computing the gradient of a given fitting function. The iterative procedure updates the solution by *moving downhill* in the residual space. The Newton–Raphson method is a one-dimensional version of gradient descent. Since it is often applied in high-dimensional settings, it is prone to find only local minima. Critical innovations for big data applications include stochastic gradient descent and the backpropagation algorithm which makes the optimization amenable to computing the gradient itself.

Alternating Descent Method (ADM) (discussed in Chapter 4) avoids computations of the gradient by optimizing in one unknown at a time. Thus all unknowns are held constant while a line search (non-convex optimization) can be performed in a single variable. This variable is then updated and held constant while another of the unknowns is updated. The iterative procedure continues through all unknowns and the iteration procedure is repeated until a desired level of accuracy is achieved.

Augmented Lagrange Method (ALM) (discussed in Chapters 3 and 8) is a class of algorithms for solving constrained optimization problems. They are similar to penalty methods in that they replace a constrained optimization problem by a series of unconstrained problems and add a penalty term to the objective which helps enforce the desired constraint. ALM adds another term designed to mimic a Lagrange multiplier. The augmented Lagrangian is not the same as the method of Lagrange multipliers.

Linear Program and Simplex Method are the workhorse algorithms for convex optimization. A linear program has an objective function which is linear in the unknown and the constraints consist of linear inequalities and equalities. By computing its feasible region, which is a convex polytope, the linear programming algorithm finds a point in the polyhedron where this function has the smallest (or largest) value if such a point exists. The simplex method is a specific iterative technique for linear programs which aims to take a given basic feasible solution to another basic feasible solution for which the objective function is smaller, thus producing an iterative procedure for optimizing.

Most Common Equations and Symbols

Linear Algebra

Linear System of Equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{0.1}$$

The matrix $\mathbf{A} \in \mathbb{R}^{p \times n}$ and vector $\mathbf{b} \in \mathbb{R}^p$ are generally known, and the vector $\mathbf{x} \in \mathbb{R}^n$ is unknown.

Eigenvalue Equation

$$\mathbf{A}\mathbf{T} = \mathbf{T}\boldsymbol{\Lambda}. \tag{0.2}$$

The columns $\boldsymbol{\xi}_k$ of the matrix $\mathbf{T}$ are the eigenvectors of $\mathbf{A} \in \mathbb{C}^{n \times n}$ corresponding to the eigenvalue λ_k: $\mathbf{A}\boldsymbol{\xi}_k = \lambda_k \boldsymbol{\xi}_k$. The matrix $\boldsymbol{\Lambda}$ is a diagonal matrix containing these eigenvalues, in the simple case with n distinct eigenvalues.

Change of Coordinates

$$\mathbf{x} = \boldsymbol{\Psi}\mathbf{a}. \tag{0.3}$$

The vector $\mathbf{x} \in \mathbb{R}^n$ may be written as $\mathbf{a} \in \mathbb{R}^n$ in the coordinate system given by the columns of $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$.

Measurement Equation

$$\mathbf{y} = \mathbf{C}\mathbf{x}. \tag{0.4}$$

The vector $\mathbf{y} \in \mathbb{R}^p$ is a measurement of the state $\mathbf{x} \in \mathbb{R}^n$ by the measurement matrix $\mathbf{C} \in \mathbb{R}^{p \times n}$.

Singular Value Decomposition

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^* \approx \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^*. \tag{0.5}$$

The matrix $\mathbf{X} \in \mathbb{C}^{n \times m}$ may be decomposed into the product of three matrices $\mathbf{U} \in \mathbb{C}^{n \times n}$, $\boldsymbol{\Sigma} \in \mathbb{C}^{n \times m}$, and $\mathbf{V} \in \mathbb{C}^{m \times m}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ are *unitary*, so that $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}_{n \times n}$ and $\mathbf{V}\mathbf{V}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I}_{m \times m}$, where * denotes complex conjugate transpose. The columns of $\mathbf{U}$ (resp. $\mathbf{V}$) are orthogonal, called left (resp. right) *singular vectors*. The matrix $\boldsymbol{\Sigma}$ contains decreasing, nonnegative diagonal entries called *singular values*.

Often, $\mathbf{X}$ is approximated with a low-rank matrix $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^*$, where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ contain the first $r \ll n$ columns of $\mathbf{U}$ and $\mathbf{V}$, respectively, and $\tilde{\boldsymbol{\Sigma}}$ contains the first $r \times r$ block of $\boldsymbol{\Sigma}$. The matrix $\tilde{\mathbf{U}}$ is often denoted $\boldsymbol{\Psi}$ in the context of spatial modes, reduced order models, and sensor placement.

Regression and Optimization

Overdetermined and Underdetermined Optimization for Linear Systems

$$\underset{\mathbf{x}}{\operatorname{argmin}} \left(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 + \lambda g(\mathbf{x}) \right) \qquad \text{or} \tag{0.6a}$$

$$\underset{\mathbf{x}}{\operatorname{argmin}}\, g(\mathbf{x}) \;\text{ subject to }\; \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \epsilon\,, \tag{0.6b}$$

Here $g(\mathbf{x})$ is a regression penalty (with penalty parameter λ for overdetermined systems). For over- and underdetermined linear systems of equations, which result in either no solutions or an infinite number of solutions of $\mathbf{A}\mathbf{x} = \mathbf{b}$, a choice of constraint or penalty, which is also known as *regularization*, must be made in order to produce a solution.

Overdetermined and Underdetermined Optimization for Nonlinear Systems

$$\underset{\mathbf{x}}{\operatorname{argmin}} \left(f(\mathbf{A}, \mathbf{x}, \mathbf{b}) + \lambda g(\mathbf{x}) \right) \qquad \text{or} \tag{0.7a}$$

$$\underset{\mathbf{x}}{\operatorname{argmin}}\, g(\mathbf{x}) \;\text{ subject to }\; f(\mathbf{A}, \mathbf{x}, \mathbf{b}) \leq \epsilon \tag{0.7b}$$

This generalizes the linear system to a nonlinear system $f(\cdot)$ with regularization $g(\cdot)$. These over- and underdetermined systems are often solved using gradient descent algorithms.

Compositional Optimization for Neural Networks

$$\underset{\mathbf{A}_j}{\operatorname{argmin}} \left(f_M(\mathbf{A}_M, \cdots f_2(\mathbf{A}_2, (f_1(\mathbf{A}_1, \mathbf{x})) \cdots) + \lambda g(\mathbf{A}_j) \right) \tag{0.8}$$

Each $\mathbf{A}_k$ denotes the weights connecting the neural network from the kth to $(k+1)$th layer. It is typically a massively underdetermined system which is regularized by $g(\mathbf{A}_j)$. Composition and regularization are critical for generating expressive representations of the data as well as preventing overfitting.

Dynamical Systems and Reduced Order Models

Nonlinear Ordinary Differential Equation (Dynamical System)

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t; \boldsymbol{\beta}). \tag{0.9}$$

The vector $\mathbf{x}(t) \in \mathbb{R}^n$ is the state of the system evolving in time t, $\boldsymbol{\beta}$ are parameters, and $\mathbf{f}$ is the vector field. Generally, $\mathbf{f}$ is Lipschitz continuous to guarantee existence and uniqueness of solutions.

Linear Input–Output System

$$\frac{d}{dt}\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{0.10a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}. \tag{0.10b}$$

The state of the system is $\mathbf{x} \in \mathbb{R}^n$, the inputs (actuators) are $\mathbf{u} \in \mathbb{R}^q$, and the outputs (sensors) are $\mathbf{y} \in \mathbb{R}^p$. The matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ define the dynamics, the effect of actuation, the sensing strategy, and the effect of actuation feed-through, respectively.

Nonlinear Map (Discrete-Time Dynamical System)

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k). \tag{0.11}$$

The state of the system at the kth iteration is $\mathbf{x}_k \in \mathbb{R}^n$, and $\mathbf{F}$ is a possibly nonlinear mapping. Often, this map defines an iteration forward in time, so that $\mathbf{x}_k = \mathbf{x}(k\Delta t)$; in this case the flow map is denoted $\mathbf{F}_{\Delta t}$.

Koopman Operator Equation (Discrete-Time)

$$\mathcal{K}_t g = g \circ \mathbf{F}_t \quad \Longrightarrow \quad \mathcal{K}_t \varphi = \lambda \varphi. \tag{0.12}$$

The linear Koopman operator $\mathcal{K}_t$ advances measurement functions of the state $g(\mathbf{x})$ with the flow $\mathbf{F}_t$. Eigenvalues and eigenvectors of $\mathcal{K}_t$ are λ and $\varphi(\mathbf{x})$, respectively. The operator $\mathcal{K}_t$ operates on a Hilbert space of measurements.

Nonlinear Partial Differential Equation

$$\mathbf{u}_t = \mathbf{N}(\mathbf{u}, \mathbf{u}_x, \mathbf{u}_{xx}, \cdots, x, t; \boldsymbol{\beta}). \tag{0.13}$$

The state of the PDE is $\mathbf{u}$, the nonlinear evolution operator is $\mathbf{N}$, subscripts denote partial differentiation, and x and t are the spatial and temporal variables, respectively. The PDE is parameterized by values in $\boldsymbol{\beta}$. The state $\mathbf{u}$ of the PDE may be a continuous function $u(x,t)$, or it may be discretized at several spatial locations, $\mathbf{u}(t) = \begin{bmatrix} u(x_1,t) & u(x_2,t) & \cdots & u(x_n,t) \end{bmatrix}^T \in \mathbb{R}^n$.

Galerkin Expansion
The continuous Galerkin expansion is:

$$u(x,t) \approx \sum_{k=1}^{r} a_k(t)\psi_k(x). \tag{0.14}$$

The functions $a_k(t)$ are temporal coefficients that capture the time dynamics, and $\psi_k(x)$ are spatial modes. For a high-dimensional discretized state, the Galerkin expansion becomes: $\mathbf{u}(t) \approx \sum_{k=1}^{r} a_k(t)\boldsymbol{\psi}_k$. The spatial modes $\boldsymbol{\psi}_k \in \mathbb{R}^n$ may be the columns of $\boldsymbol{\Psi} = \tilde{\mathbf{U}}$.

Complete Symbols

Dimensions

K	Number of nonzero entries in a K-sparse vector $\mathbf{s}$
m	Number of data snapshots (i.e., columns of $\mathbf{X}$)
n	Dimension of the state, $\mathbf{x} \in \mathbb{R}^n$
p	Dimension of the measurement or output variable, $\mathbf{y} \in \mathbb{R}^p$
q	Dimension of the input variable, $\mathbf{u} \in \mathbb{R}^q$
r	Rank of truncated SVD, or other low-rank approximation

Scalars

s	Frequency in Laplace domain
t	Time
δ	learning rate in gradient descent
Δt	Time step
x	Spatial variable
Δx	Spatial step
σ	Singular value
λ	Eigenvalue
λ	Sparsity parameter for sparse optimization (Section 7.3)
λ	Lagrange multiplier (Sections. 3.7, 8.4, and 11.4)
τ	Threshold

Vectors

$\mathbf{a}$	Vector of mode amplitudes of $\mathbf{x}$ in basis $\boldsymbol{\Psi}$, $\mathbf{a} \in \mathbb{R}^r$
$\mathbf{b}$	Vector of measurements in linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$
$\mathbf{b}$	Vector of DMD mode amplitudes (Section 7.2)
$\mathbf{Q}$	Vector containing potential function for PDE-FIND
$\mathbf{r}$	Residual error vector
$\mathbf{s}$	Sparse vector, $\mathbf{s} \in \mathbb{R}^n$
$\mathbf{u}$	Control variable (Chapters 8, 9, and 10)
$\mathbf{u}$	PDE state vector (Chapters 11 and 12)
$\mathbf{w}$	Exogenous inputs
$\mathbf{w}_d$	Disturbances to system
$\mathbf{w}_n$	Measurement noise
$\mathbf{w}_r$	Reference to track
$\mathbf{x}$	State of a system, $\mathbf{x} \in \mathbb{R}^n$
$\mathbf{x}_k$	Snapshot of data at time t_k
$\mathbf{x}_j$	Data sample $j \in Z := \{1, 2, \cdots, m\}$ (Chapters 5 and 6)
$\tilde{\mathbf{x}}$	Reduced state, $\tilde{\mathbf{x}} \in \mathbb{R}^r$, so that $\mathbf{x} \approx \tilde{\mathbf{U}}\tilde{\mathbf{x}}$
$\hat{\mathbf{x}}$	Estimated state of a system
$\mathbf{y}$	Vector of measurements, $\mathbf{y} \in \mathbb{R}^p$
$\mathbf{y}_j$	Data label $j \in Z := \{1, 2, \cdots, m\}$ (Chapters 5 and 6)
$\hat{\mathbf{y}}$	Estimated output measurement
$\mathbf{z}$	Transformed state, $\mathbf{x} = \mathbf{T}\mathbf{z}$ (Chapters 8 and 9)
$\boldsymbol{\epsilon}$	Error vector

Vectors, continued

- $\boldsymbol{\beta}$ Bifurcation parameters
- $\boldsymbol{\xi}$ Eigenvector of Koopman operator (Sections 7.4 and 7.5)
- $\boldsymbol{\xi}$ Sparse vector of coefficients (Section 7.3)
- $\boldsymbol{\phi}$ DMD mode
- $\boldsymbol{\psi}$ POD mode
- $\boldsymbol{\Upsilon}$ Vector of PDE measurements for PDE-FIND

Matrices

- $\mathbf{A}$ Matrix for system of equations or dynamics
- $\tilde{\mathbf{A}}$ Reduced dynamics on r-dimensional POD subspace
- $\mathbf{A}_\mathbf{X}$ Matrix representation of linear dynamics on the state $\mathbf{x}$
- $\mathbf{A}_\mathbf{Y}$ Matrix representation of linear dynamics on the observables $\mathbf{y}$
- $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{B})$ Matrices for continuous-time state-space system
- $(\mathbf{A}_d, \mathbf{B}_d, \mathbf{C}_d, \mathbf{B}_d)$ Matrices for discrete-time state-space system
- $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{B}})$ Matrices for state-space system in new coordinates $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$
- $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \tilde{\mathbf{B}})$ Matrices for reduced state-space system with rank r
- $\mathbf{B}$ Actuation input matrix
- $\mathbf{C}$ Linear measurement matrix from state to measurements
- $\mathcal{C}$ Controllability matrix
- $\mathcal{F}$ Discrete Fourier transform
- $\mathbf{G}$ Matrix representation of linear dynamics on the states and inputs $[\mathbf{x}^T \mathbf{u}^T]^T$
- $\mathbf{H}$ Hankel matrix
- $\mathbf{H}'$ Time-shifted Hankel matrix
- $\mathbf{I}$ Identity matrix
- $\mathbf{K}$ Matrix form of Koopman operator (Chapter 7)
- $\mathbf{K}$ Closed-loop control gain (Chapter 8)
- $\mathbf{K}_f$ Kalman filter estimator gain
- $\mathbf{K}_r$ LQR control gain
- $\mathbf{L}$ Low-rank portion of matrix $\mathbf{X}$ (Chapter 3)
- $\mathcal{O}$ Observability matrix
- $\mathbf{P}$ Unitary matrix that acts on columns of $\mathbf{X}$
- $\mathbf{Q}$ Weight matrix for state penalty in LQR (Sec. 8.4)
- $\mathbf{Q}$ Orthogonal matrix from QR factorization
- $\mathbf{R}$ Weight matrix for actuation penalty in LQR (Sec. 8.4)
- $\mathbf{R}$ Upper triangular matrix from QR factorization
- $\mathbf{S}$ Sparse portion of matrix $\mathbf{X}$ (Chapter 3)
- $\mathbf{T}$ Matrix of eigenvectors (Chapter 8)
- $\mathbf{T}$ Change of coordinates (Chapters 8 and 9)
- $\mathbf{U}$ Left singular vectors of $\mathbf{X}$, $\mathbf{U} \in \mathbb{R}^{n\times n}$
- $\hat{\mathbf{U}}$ Left singular vectors of economy SVD of $\mathbf{X}$, $\mathbf{U} \in \mathbb{R}^{n\times m}$
- $\tilde{\mathbf{U}}$ Left singular vectors (POD modes) of truncated SVD of $\mathbf{X}$, $\mathbf{U} \in \mathbb{R}^{n\times r}$
- $\mathbf{V}$ Right singular vectors of $\mathbf{X}$, $\mathbf{V} \in \mathbb{R}^{m\times m}$
- $\tilde{\mathbf{V}}$ Right singular vectors of truncated SVD of $\mathbf{X}$, $\mathbf{V} \in \mathbb{R}^{m\times r}$

Matrices, continued

$\mathbf{\Sigma}$	Matrix of singular values of $\mathbf{X}$, $\mathbf{\Sigma} \in \mathbb{R}^{n\times m}$
$\hat{\mathbf{\Sigma}}$	Matrix of singular values of economy SVD of $\mathbf{X}$, $\mathbf{\Sigma} \in \mathbb{R}^{m\times m}$
$\tilde{\mathbf{\Sigma}}$	Matrix of singular values of truncated SVD of $\mathbf{X}$, $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$
$\mathbf{W}$	Eigenvectors of $\tilde{\mathbf{A}}$
$\mathbf{W}_c$	Controllability Gramian
$\mathbf{W}_o$	Observability Gramian
$\mathbf{X}$	Data matrix, $\mathbf{X} \in \mathbb{R}^{n\times m}$
$\mathbf{X}'$	Time-shifted data matrix, $\mathbf{X}' \in \mathbb{R}^{n\times m}$
$\mathbf{Y}$	Projection of $\mathbf{X}$ matrix onto orthogonal basis in randomized SVD (Sec. 1.8)
$\mathbf{Y}$	Data matrix of observables, $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, $\mathbf{Y} \in \mathbb{R}^{p\times m}$ (Chapter 7)
$\mathbf{Y}'$	Shifted data matrix of observables, $\mathbf{Y}' = \mathbf{g}(\mathbf{X}')$, $\mathbf{Y}' \in \mathbb{R}^{p\times m}$ (Chapter 7)
$\mathbf{Z}$	Sketch matrix for randomized SVD, $\mathbf{Z} \in \mathbb{R}^{n\times r}$ (Sec. 1.8)
$\mathbf{\Theta}$	Measurement matrix times sparsifying basis, $\mathbf{\Theta} = \mathbf{C}\mathbf{\Psi}$ (Chapter 3)
$\mathbf{\Theta}$	Matrix of candidate functions for SINDy (Sec. 7.3)
$\mathbf{\Gamma}$	Matrix of derivatives of candidate functions for SINDy (Sec. 7.3)
$\mathbf{\Xi}$	Matrix of coefficients of candidate functions for SINDy (Sec. 7.3)
$\mathbf{\Xi}$	Matrix of nonlinear snapshots for DEIM (Sec. 12.5)
$\mathbf{\Lambda}$	Diagonal matrix of eigenvalues
$\mathbf{\Upsilon}$	Input snapshot matrix, $\mathbf{\Upsilon} \in \mathbb{R}^{q\times m}$
$\mathbf{\Phi}$	Matrix of DMD modes, $\mathbf{\Phi} \triangleq \mathbf{X}'\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}$
$\mathbf{\Psi}$	Orthonormal basis (e.g., Fourier or POD modes)

Tensors

$(\mathcal{A}, \mathcal{B}, \mathcal{M})$	N-way array tensors of size $I_1 \times I_2 \times \cdots \times I_N$

Norms

$\|\cdot\|_0$	ℓ_0 pseudo-norm of a vector $\mathbf{x}$ the number of nonzero elements in $\mathbf{x}$
$\|\cdot\|_1$	ℓ_1 norm of a vector $\mathbf{x}$ given by $\|\mathbf{x}\|_1 = \sum_{i=1}^n \lvert x_i\rvert$
$\|\cdot\|_2$	ℓ_2 norm of a vector $\mathbf{x}$ given by $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n (x_i^2)}$
$\|\cdot\|_2$	2-norm of a matrix $\mathbf{X}$ given by $\|\mathbf{X}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{X}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$
$\|\cdot\|_F$	Frobenius norm of a matrix $\mathbf{X}$ given by $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m \lvert X_{ij}\rvert^2}$
$\|\cdot\|_*$	Nuclear norm of a matrix $\mathbf{X}$ given by $\|\mathbf{X}\|_* = \text{trace}\left(\sqrt{\mathbf{X}^*\mathbf{X}}\right) = \sum_{i=1}^m \sigma_i$ (for $m \leq n$)
$\langle\cdot,\cdot\rangle$	Inner product. For functions, $\langle f(x), g(x)\rangle = \int_{-\infty}^{\infty} f(x)g^*(x)dx$.
$\langle\cdot,\cdot\rangle$	Inner product. For vectors, $\langle\mathbf{u},\mathbf{v}\rangle = \mathbf{u}^*\mathbf{v}$.

Operators, Functions, and Maps

$\mathcal{F}$	Fourier transform
$\mathbf{F}$	Discrete-time dynamical system map
$\mathbf{F}_t$	Discrete-time flow map of dynamical system through time t
$\mathbf{f}$	Continuous-time dynamical system
$\mathcal{G}$	Gabor transform

Operators, Functions, and Maps, continued

$\mathbf{G}$ Transfer function from inputs to outputs (Chapter 8)
g Scalar measurement function on $\mathbf{x}$
$\mathbf{g}$ Vector-valued measurement functions on $\mathbf{x}$
J Cost function for control
ℓ Loss function for support vector machines (Chapter 5)
$\mathcal{K}$ Koopman operator (continuous time)
$\mathcal{K}_t$ Koopman operator associated with time t flow map
$\mathcal{L}$ Laplace transform
$\mathbf{L}$ Loop transfer function (Chapter 8)
$\mathbf{L}$ Linear partial differential equation (Chapters 11 and 12)
$\mathbf{N}$ Nonlinear partial differential equation
$\mathcal{O}$ Order of magnitude
$\mathbf{S}$ Sensitivity function (Chapter 8)
$\mathbf{T}$ Complementary sensitivity function (Chapter 8)
$\mathcal{W}$ Wavelet transform
μ Incoherence between measurement matrix $\mathbf{C}$ and basis $\boldsymbol{\Psi}$
κ Condition number
φ Koopman eigenfunction
∇ Gradient operator
$*$ Convolution operator

Most Common Acronyms

CNN	Convolutional neural network
DL	Deep learning
DMD	Dynamic mode decomposition
FFT	Fast Fourier transform
ODE	Ordinary differential equation
PCA	Principal components analysis
PDE	Partial differential equation
POD	Proper orthogonal decomposition
ROM	Reduced order model
SVD	Singular value decomposition

Other Acronyms

ADM	Alternating directions method
AIC	Akaike information criterion
ALM	Augmented Lagrange multiplier
ANN	Artificial neural network
ARMA	Autoregressive moving average
ARMAX	Autoregressive moving average with exogenous input
BIC	Bayesian information criterion
BPOD	Balanced proper orthogonal decomposition
DMDc	Dynamic mode decomposition with control
CCA	Canonical correlation analysis
CFD	Computational fluid dynamics
CoSaMP	Compressive sampling matching pursuit
CWT	Continuous wavelet transform
DEIM	Discrete empirical interpolation method
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
DMDc	Dynamic mode decomposition with control
DNS	Direct numerical simulation
DWT	Discrete wavelet transform
ECOG	Electrocorticography
eDMD	Extended DMD
EIM	Empirical interpolation method
EM	Expectation maximization
EOF	Empirical orthogonal functions
ERA	Eigensystem realization algorithm
ESC	Extremum-seeking control
GMM	Gaussian mixture model
HAVOK	Hankel alternative view of Koopman
JL	Johnson–Lindenstrauss
KL	Kullback–Leibler
ICA	Independent component analysis

Other Acronyms, continued

KLT	Karhunen–Loève transform
LAD	Least absolute deviations
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
LQE	Linear quadratic estimator
LQG	Linear quadratic Gaussian controller
LQR	Linear quadratic regulator
LTI	Linear time invariant system
MIMO	Multiple input, multiple output
MLC	Machine learning control
MPE	Missing point estimation
mrDMD	Multi-resolution dynamic mode decomposition
NARMAX	Nonlinear autoregressive model with exogenous inputs
NLS	Nonlinear Schrödinger equation
OKID	Observer Kalman filter identification
PBH	Popov–Belevitch–Hautus test
PCP	Principal component pursuit
PDE-FIND	Partial differential equation functional identification of nonlinear dynamics
PDF	Probability distribution function
PID	Proportional-integral-derivative control
PIV	Particle image velocimetry
RIP	Restricted isometry property
rSVD	Randomized SVD
RKHS	Reproducing kernel Hilbert space
RNN	Recurrent neural network
RPCA	Robust principal components analysis
SGD	Stochastic gradient descent
SINDy	Sparse identification of nonlinear dynamics
SISO	Single input, single output
SRC	Sparse representation for classification
SSA	Singular spectrum analysis
STFT	Short time Fourier transform
STLS	Sequential thresholded least-squares
SVM	Support vector machine
TICA	Time-lagged independent component analysis
VAC	Variational approach of conformation dynamics