

Matrix Methods in Signal Processing ...

(Lecture notes for EECS 551)

Jeff Fessler
University of Michigan

June 18, 2020

Contents

0	EECS 551 Course introduction: F19	0.1
	0.1 Course logistics	0.2
	0.2 Julia language	0.12
	0.3 Course topics	0.19
1	Introduction to Matrices	1.1
	1.0 Introduction	1.2
	1.1 Basics	1.3
	1.2 Matrix structures	1.13
	Notation	1.13
	Common matrix shapes and types	1.14
	Matrix transpose and symmetry	1.19

1.3 Multiplication	1.21
Vector-vector multiplication	1.21
Matrix-vector multiplication	1.24
Matrix-matrix multiplication	1.30
Matrix multiplication properties	1.31
Kronecker product and Hadamard product and the vec operator	1.37
Using matrix-vector operations in high-level computing languages	1.39
Invertibility	1.47
1.4 Orthogonality	1.51
Orthogonal vectors	1.51
Cauchy-Schwarz inequality	1.53
Orthogonal matrices	1.54
1.5 Determinant of a matrix	1.56
1.6 Eigenvalues	1.64
Properties of eigenvalues	1.68
1.7 Trace	1.71
1.8 Appendix: Fields, Vector Spaces, Linear Transformations	1.72
2 Matrix factorizations / decompositions	2.1
2.0 Introduction	2.2
Matrix factorizations	2.3

2.1 Spectral Theorem (for symmetric matrices)	2.5
Normal matrices	2.7
Square asymmetric and non-normal matrices	2.10
Geometry of matrix diagonalization	2.12
2.2 SVD	2.20
Existence of SVD	2.21
Geometry	2.22
2.3 The matrix 2-norm or spectral norm	2.27
Eigenvalues as optimization problems	2.31
2.4 Relating SVDs and eigendecompositions	2.32
When does $U = V$?	2.36
2.5 Positive semidefinite matrices	2.40
2.6 Summary	2.43
SVD computation using eigendecomposition	2.44
3 Subspaces and rank	3.1
3.0 Introduction	3.3
3.1 Subspaces	3.4
Span	3.7
Linear independence	3.10
Basis	3.12

Dimension	3.16
Sums and intersections of subspaces	3.17
Direct sum of subspaces	3.19
Dimensions of sums of subspaces	3.20
Orthogonal complement of a subspace	3.21
Linear transformations	3.22
Range of a matrix	3.23
3.2 Rank of a matrix	3.25
Rank of a matrix product	3.28
Unitary invariance of rank / eigenvalues / singular values	3.31
3.3 Nullspace and the SVD	3.33
Nullspace or kernel	3.33
The four fundamental spaces	3.37
Anatomy of the SVD	3.39
SVD of finite differences (discrete derivative)	3.43
Synthesis view of matrix decomposition	3.46
3.4 Orthogonal bases	3.47
3.5 Spotting eigenvectors	3.51
3.6 Application: Signal classification by nearest subspace	3.55
Projection onto a set	3.55
Nearest point in a subspace	3.56

Optimization preview	3.58
3.7 Summary	3.60
4 Linear equations and least-squares	4.1
4.0 Introduction to linear equations	4.2
Linear regression and machine learning	4.4
4.1 Linear least-squares estimation	4.6
Minimization and gradients	4.10
Solving LLS using the normal equations	4.15
Solving LLS problems using the compact SVD	4.16
Uniqueness of LLS solution	4.21
Moore-Penrose pseudoinverse	4.23
4.2 Linear least-squares estimation: Under-determined case	4.30
Orthogonality principle	4.32
Minimum-norm LS solution via pseudo-inverse	4.35
4.3 Truncated SVD solution	4.39
Low-rank approximation interpretation of truncated SVD solution	4.42
Noise effects	4.44
Tikhonov regularization aka ridge regression	4.46
4.4 Summary of LLS solution methods in terms of SVD	4.48
4.5 Frames and tight frames	4.49

4.6 Projection and orthogonal projection	4.55
Projection onto a subspace	4.61
Binary classifier design using least-squares	4.69
4.7 Summary	4.71
5 Norms	5.1
5.0 Introduction	5.2
5.1 Vector norms	5.3
Properties of norms	5.7
Norm notation	5.9
Unitarily invariant norms	5.10
Inner products	5.11
5.2 Matrix norms and operator norms	5.17
Induced matrix norms	5.21
Norms defined in terms of singular values	5.24
Properties of matrix norms	5.27
Spectral radius	5.30
5.3 Convergence of sequences of vectors and matrices	5.35
5.4 Generalized inverse of a matrix	5.37
5.5 Procrustes analysis	5.39
Generalizations: non-square, complex, with translation	5.46

5.6 Summary	5.51
6 Low-rank approximation	6.1
6.0 Introduction	6.2
6.1 Low-rank approximation via Frobenius norm	6.3
Implementation	6.8
1D example	6.15
Generalization to other norms	6.17
Bases for $\mathbb{F}^{M \times N}$	6.19
Low-rank approximation summary	6.22
Rank and stability	6.23
6.2 Sensor localization application (Multidimensional scaling)	6.24
Practical implementation	6.31
6.3 Proximal operators	6.34
6.4 Alternative low-rank approximation formulations	6.38
6.5 Choosing the rank or regularization parameter	6.46
OptShrink	6.50
6.6 Related methods: autoencoders and PCA	6.55
Relation to autoencoder with linear hidden layer	6.55
Relation to principal component analysis (PCA)	6.58
6.7 Subspace learning	6.60

6.8 Summary	6.65
7 Special matrices	7.1
7.0 Introduction	7.2
7.1 Companion matrices	7.2
Vandermonde matrices and diagonalizing a companion matrix	7.9
Using companion matrices to check for common roots of two polynomials	7.11
7.2 Circulant matrices	7.13
7.3 Toeplitz matrices	7.20
7.4 Power iteration	7.23
Geršgorin disk theorem	7.28
7.5 Nonnegative matrices and Perron-Frobenius theorem	7.31
Markov chains	7.36
Irreducible matrix	7.46
Google's PageRank method	7.55
7.6 Summary	7.59
8 Optimization basics	8.1
8.0 Introduction	8.2
8.1 Preconditioned gradient descent (PGD) for LS	8.3
Tool: Matrix square root	8.4

Convergence rate analysis of PGD: first steps	8.8
Tool: Matrix powers	8.9
Classical GD: step size bounds	8.11
Optimal step size for GD	8.12
Practical step size for GD	8.13
Ideal preconditioner for PGD	8.14
Tool: Positive (semi)definiteness properties	8.15
General preconditioners for PGD	8.17
Diagonal majorizer	8.18
Tool: commuting (square) matrices	8.24
8.2 Preconditioned steepest descent	8.28
8.3 Gradient descent for smooth convex functions	8.29
8.4 Machine learning via logistic regression for binary classification	8.36
8.5 Summary	8.44
9 Matrix completion	9.1
9.0 Introduction	9.2
9.1 Measurement model	9.3
9.2 LRMC: noiseless case	9.7
Noiseless problem statement	9.7
Alternating projection approach to LRMC	9.8

9.3 LRMC: noisy case	9.13
Noisy problem statement	9.13
Majorize-minimize (MM) iterations	9.15
MM methods for LRMC	9.16
LRMC by iterative low-rank approximation	9.18
LRMC by iterative singular value hard thresholding	9.19
LRMC by iterative singular value soft thresholding (ISTA)	9.20
Demo	9.23
9.4 Summary	9.24
99 Miscellaneous topics / review	99.1
99.0 Review / practice questions	99.2
Ch01: Matrices	99.2
Ch02: Matrix decompositions	99.3
Ch03: Subspaces	99.5
Ch04: Linear least-squares	99.7
Ch05: Norms	99.10
Ch06: Low-rank approximation	99.11
Ch07: Special matrices	99.13
Ch08: Optimization	99.15
Ch09: Matrix completion	99.17

Index

- ℓ_p norm, 5.3
- $\ell_{p,q}$ matrix norms, 5.16
- n -tuple space, 1.69
- (Eckart-Young-Mirsky), 6.16
- (matrix) square root, 8.3
- (proof), 1.56
- compact SVD, 2.23, 3.37, 4.12, 4.18, 4.20, 4.22, 4.59
- “dangerous bend” symbol, 0.8
- 2 norm, 1.47
- 802.11n wifi, 2.26

- absorbing, 7.39
- absorbing states, 7.41
- additive synthesis, 3.13
- additivity, 5.8

- adjacency matrix, 1.9
- affine, 6.53
- algebraic multiplicity, 7.24, 7.36
- alternating projection, 9.8
- angle, 1.48, 5.11
- ANN, 6.52
- aperiodic, 7.35, 7.36
- array multiplication, 9.7
- artificial neural network, 6.52
- associative property, 1.29
- asymptotics, 7.32
- autoencoder, 6.52

- basis, 3.11, 3.13, 3.14, 3.43, 6.18, 6.19
- BCCB, 1.18

- beamforming, 2.27
- bias vector, 6.53
- block circulant matrix, 1.18
- block circulant with circulant blocks, 1.18
- block diagonal, 7.4
- block diagonal matrix, 1.18
- block lower triangular, 1.55
- block matrix, 1.18
- block matrix multiplication, 1.34
- block matrix triangularization, 1.55
- block upper triangular, 1.55
- bounded curvature, 8.26
- broadcast, 1.38

- call by reference, 0.12
- canonical basis, 3.44
- Cauchy-Bunyakovsky-Schwarz, 1.48, 5.10
- Cauchy-Schwarz inequality, 1.48, 5.10
- channel estimation, 2.27
- characteristic equation, 1.58
- characteristic polynomial, 1.58, 7.2, 7.5
- Cholesky decomposition, 1.14, 2.2

- Circulant, 1.16
- circulant, 7.34, 8.21
- circulant matrix, 2.7, 7.10, 7.12, 7.15
- Cleve Moler, 7.43
- CNN, 1.6
- code, 6.52
- coil compression, 6.11
- column rank, 3.23
- column space, 3.21
- commutative property, 1.29, 1.63
- commute, 7.14
- commuting matrices, 8.1, 8.21
- compact SVD, 2.19, 4.23, 4.26, 4.44, 4.60, 6.27, 6.28
- companion, 8.21
- companion matrices, 7.6
- companion matrix, 7.2, 7.5, 7.10
- compatible, 5.15
- complements, 3.17
- completed the square, 4.12
- Complex Euclidean n -dimensional space, 1.69
- composite cost function, 9.20

- compressed sensing, 4.32
- concatenation, 1.29
- condition number, 4.36
- conjugate gradient method, 4.42
- conjugate transpose, 1.19, 1.20
- consistent, 5.13
- constrained optimization, 8.31
- convergence, 5.45
- converges, 5.28–5.30, 7.24, 8.1, 8.29
- convex, 6.38, 8.25, 8.27, 8.29, 9.8, 9.19
- convex combination, 8.30
- convex function, 4.11, 5.6
- convex hull, 3.7
- convex relaxation, 6.38, 9.13
- convex set, 8.30
- convolution, 1.6
- convolutional neural network, 1.6
- coordinate system, 3.11
- coordinates, 3.11
- correlation coefficient, 5.11
- cost function, 8.2
- counting measure, 5.4
- covariance matrix, 1.59
- cyclic commutative property, 1.65
- damping factor, 7.42
- data, 1.4
- DCT, 3.11, 3.40
- de-mean, 6.25
- decoder, 6.52
- decomposition methods, 1.55
- decrease monotonically, 8.22
- degree, 1.58
- degrees of freedom, 2.10, 9.5
- dense matrix, 1.16
- derivative, 3.39, 4.10
- determinant, 1.51, 1.52, 7.3
- determinant commutative property, 1.52
- DFT, 1.4, 2.7, 7.11
- DFT matrix, 7.7
- diagonal, 1.14, 2.10, 5.30, 7.20
- diagonalizable, 2.6, 2.9, 2.10, 2.16, 7.7, 7.16, 8.4, 8.7
- diagonally preconditioned GD, 8.16

- dimension, 3.14
- dimensionality reduction, 3.3, 6.1, 6.2, 6.14, 6.18, 6.52
- direct sum, 3.17
- directed graph, 1.9, 7.26
- discrete cosine transform, 3.11, 3.40
- discrete sine transform, 3.40
- displacement vector, 5.40
- distance matrix, 6.23
- distributive property, 1.29, 1.62
- Divergence, 6.45
- DoF, 2.10, 9.5
- dot product, 1.21, 1.40
- DST, 3.40

- economy SVD, 4.16, 4.44
- eigendecomposition, 2.2–2.4, 8.7
- eigendecompositions, 2.10
- eigenfunctions, 3.39
- eigenvalue, 1.9, 7.18
- eigenvalue algorithms, 1.15
- eigenvalues, 1.58, 1.63, 2.3, 7.2, 7.5

- eigenvector, 1.9, 1.59, 2.3, 2.34
- eigenvectors, 2.4, 7.6
- element-wise, 1.36
- EM algorithm, 9.22
- empirical covariance matrix, 6.55
- encoder, 6.52
- equilibrium, 7.30
- equilibrium distribution, 7.29, 7.31, 7.39
- equivalent, 5.24
- error, 4.29
- Euclidean n -dimensional space, 1.69
- Euclidean norm, 1.50, 5.2
- Euclidian norm, 5.3
- exists, 7.29

- factor analysis, 6.7
- Fair potential, 8.28
- fast Fourier transform, 7.13, 7.15
- FEM, 3.40
- FFT, 7.13, 7.15
- field, 1.67
- field of scalars, 1.67

- Finite difference, 1.14
- finite difference, 3.39
- finite-dimensional, 3.14
- finite-element method, 3.40
- FISTA, 9.22
- flat, 4.30
- floating-point operations, 1.24
- floating-point precision, 4.35
- FLOPs, 1.24
- FOIL, 5.35
- Fourier series, 3.6
- frame, 4.43, 4.45
- frame bounds, 4.43
- frames, 7.7
- Frobenius inner product, 5.9, 5.12, 5.15
- Frobenius matrix norm, 5.33
- Frobenius norm, 5.14, 5.25, 5.32, 5.45, 6.2, 6.5
- full matrix, 1.16
- full SVD, 4.18, 4.23
- fundamental theorem of algebra, 1.58
- fundamental theorem of linear algebra, 3.33
- Gaussian elimination, 1.14, 2.2, 4.11
- GD, 5.28, 8.2
- Gelfand's formula, 5.27, 7.23
- generalized inverse, 5.31, 5.45
- generalized principal component analysis, 6.59
- geodesic distances, 6.32
- geometric multiplicity, 7.24, 7.36
- Geršgorin disk theorem, 7.22, 7.23
- Google's PageRank, 7.40
- GP, 8.31
- GPCA, 6.59
- gradient, 4.11
- gradient descent, 5.28, 8.2, 8.25
- gradient projection, 8.31
- gradients, 8.25
- Gram matrix, 1.15, 1.51, 1.59, 4.11, 6.25, 6.26
- Gram-Schmidt, 2.2
- group sparsity, 5.16
- Hölder's inequality, 5.12
- Hadamard product, 9.7
- Heaviside step function, 6.34

- Hermitian, 1.19, 4.50
- Hermitian symmetric, 1.19
- Hermitian symmetry, 5.8
- Hermitian transpose, 1.19, 1.20
- Hessian matrix, 8.26, 8.33
- Hilbert–Schmidt norm, 5.14
- Hilbert–Schmidt inner product, 5.9
- hull, 3.7
- hyperparameter, 6.35

- idempotent, 4.24, 4.49, 4.50
- idempotent matrix, 4.49, 5.43
- identity, 1.67
- ill posed, 9.2
- image compression, 6.1
- image registration, 5.34
- improper rotation, 2.14
- indicator function, 5.4
- induced, 5.17
- induced matrix norm, 5.27
- induced matrix norms, 5.27
- induced norm, 5.9, 5.17, 5.22

- infinite dimensional, 3.6, 3.14
- infinity norm, 5.3
- information retrieval, 1.7
- inner product, 1.21, 1.30, 5.8, 5.9, 5.12, 6.19
- inner product spaces, 5.8
- integer programming, 6.59
- intersection, 3.15
- invariant, 5.38
- inverse, 1.67
- invertible, 1.43, 1.52, 2.9, 4.8, 7.7, 8.7
- invertible matrix, 1.43, 3.22
- irreducible, 7.34–7.37, 7.39, 7.41
- irreducible matrices, 7.24, 7.36
- irreducible matrix, 7.36
- ISTA, 9.21
- iterative shrinkage thresholding algorithm, 9.21
- iterative soft-thresholding algorithm, 9.21

- Jordan form, 7.19
- Jordan normal form, 1.65, 2.10
- Julia programming language, 0.9

- K-nearest-neighbors method, 6.56
- kernel, 3.29
- Kronecker sum, 7.8
- Ky-Fan K -norm, 5.20

- labeled, 6.59
- landmark registration, 5.34
- Laplace's formula, 1.54
- latent, 6.43, 9.3
- latent variable, 6.52
- law of total probability, 7.30
- LDA, 1.21
- left eigenvector, 1.9
- left inverse, 1.43, 4.19, 5.31
- left singular vector, 2.28, 2.34
- left singular vectors, 2.18, 2.38
- limiting behavior, 7.32
- line search, 8.24
- linear, 3.20
- linear algebra, 1.21
- linear combination, 3.9
- linear combinations, 3.8
- linear discriminant analysis, 1.21
- linear least-squares, 4.6, 5.38, 6.14, 6.15
- linear least-squares estimate, 4.6
- linear map, 1.4, 3.20
- linear mapping, 1.70
- linear operation, 1.4, 1.10, 1.11
- linear operator, 1.70
- linear regression, 4.4
- linear space, 1.68
- linear span, 3.7
- linear subspace, 3.4
- linear transform, 3.20
- linear transformation, 1.4, 1.70, 3.20
- linear variety, 4.30, 5.31
- linearly dependent, 3.9, 3.10
- linearly independent, 2.9, 3.6, 3.9, 3.10, 3.43, 4.8, 6.18
- link graph matrix, 1.9
- Lipschitz continuous, 8.25, 8.28, 8.29
- Lipschitz continuous gradient, 8.25, 8.26
- LLS, 4.6

- logistic, 8.33
- logistic regression, 4.61, 8.35
- low rank, 9.2
- low-rank, 6.11, 6.27, 9.7
- low-rank approximation, 4.38, 6.1, 6.19, 6.53
- low-rank approximation problem, 6.2, 6.16
- low-rank matrix completion, 9.2
- lower bound on the number of samples, 9.5
- lower Hessenberg, 1.15
- lower triangular, 1.14, 7.4
- LRMC, 9.2
- LTI, 1.6
- LU decomposition, 2.2

- majorization, 8.1
- majorize-minimize, 9.14
- majorizer, 8.15, 9.14, 9.15
- majorizes, 8.15
- Markov chain, 7.25
- Markov chains, 7.20
- matched filter, 1.21
- matrices, 6.18

- matrix, 1.4, 1.8, 1.9, 1.20, 6.52, 7.27
- matrix 2-norm, 2.25
- matrix addition, 8.14
- matrix completion, 6.40, 9.2
- matrix determinant lemma, 1.56
- matrix exponential, 1.36
- matrix inversion lemma, 1.44
- matrix multiplication, 1.29, 9.7
- matrix norm, 5.13, 5.26
- matrix norms, 5.2, 5.13, 5.29, 5.45
- matrix powers, 8.1, 8.7
- matrix sensing, 6.40
- matrix square root, 8.1, 8.4, 8.5
- matrix-matrix product, 1.29, 1.30
- matrix-vector, 1.33
- matrix-vector multiplication, 1.11
- matrix-vector product, 1.24–1.26
- max norm, 5.3, 5.14
- maximum column sum matrix norm, 5.19
- maximum row sum matrix norm, 5.18
- MIMO, 2.26

- minimal polynomials, 2.10
- minimum norm, 4.27
- minimum norm LLS solution, 4.31
- minimum polynomial, 7.5
- minor, 1.54
- MM, 9.14
- model, 9.2
- monic polynomial, 7.2, 7.5
- monomials, 3.10
- monotone, 8.22
- monotonic, 8.22
- Moore-Penrose pseudo-inverse, 5.45
- Moore-Penrose pseudoinverse, 4.18, 4.19, 5.31
- MSE, 6.45
- Multi-input multi-output, 2.26
- multidimensional scaling, 6.23, 6.28
- multiple dispatch, 1.39
- multiplication by a scalar, 1.68
- multiplicity, 7.23
- multiplicity one, 7.24, 7.36
- necessary condition, 4.28
- Netflix problem, 9.2
- NMF, 6.21
- non-convex, 6.2, 9.7
- nonnegative, 7.22, 7.28, 7.35
- nonnegative matrices, 7.22
- nonnegative matrix, 7.20
- nonnegative matrix factorizations, 6.21
- nonnegative orthant, 8.30
- norm, 1.47, 5.5, 5.9, 5.29
- normal, 2.6–2.9, 2.29–2.31, 2.34, 3.28, 5.26
- normal equations, 4.11, 4.24, 4.28
- normal matrix, 2.6
- normalized root mean-squared difference, 6.43
- normalized root mean-squared error, 6.43
- NP hard, 9.7
- NRMSD, 6.43
- NRMSE, 6.43
- nuclear norm, 5.20, 5.24, 6.38, 9.13, 9.19
- null space, 3.29, 4.23
- nullity, 3.32
- nullspace, 3.29

- OGM, 8.24
- one, 1.67
- operator norm, 5.17
- operator norms, 5.2, 5.13, 5.18, 5.45
- optimization, 8.1
- optimized gradient method, 8.24
- ordered weighted ℓ_1 , 5.45
- orthogonal, 1.46, 1.50, 1.62, 2.4, 2.12, 3.5, 3.43
- orthogonal basis, 3.43, 3.46
- orthogonal complement, 3.19, 4.58, 6.25
- orthogonal matrix, 1.49, 3.45, 3.46, 4.50
- orthogonal polynomials, 4.8
- orthogonal Procrustes problem, 5.33, 5.36, 5.37, 5.44, 5.45
- orthogonal projection matrix, 4.50, 4.51, 4.60
- orthogonal projector, 4.24, 4.50
- orthogonal projectors, 4.59
- orthogonal set, 1.46
- orthogonal wavelet transform, 3.11
- orthogonality principle, 4.28, 4.57
- orthonormal, 1.46, 6.54, 7.11
- orthonormal bases, 4.56, 4.59
- orthonormal basis, 2.4, 3.44, 3.50, 6.20, 6.57
- orthonormal columns, 4.20
- orthonormal rows, 4.20
- orthonormal set, 1.46, 1.49
- outer product, 1.22, 1.34, 3.27, 3.32, 3.46, 5.15
- outer-product matrix, 1.59
- outlier, 5.46
- over-determined system, 4.17
- OWL, 5.45
- OWT, 3.11
- PageRank, 1.9, 7.20
- parallel processing, 1.30
- parallelogram law, 5.9
- Parseval tight frame, 4.46–4.48
- Parseval's theorem, 1.50, 5.7, 6.6
- PCA, 6.21, 6.54
- PCA generalizations, 6.21
- pentadiagonal, 1.15
- perceptron, 1.21
- period, 7.35

- period of the i th index, 7.35
- periodic functions, 3.6
- permutation matrix, 2.7
- permutation method, 6.10
- perpendicular, 1.46, 4.28
- Perron root, 7.22, 7.24, 7.36
- Perron vector, 7.24
- Perron vectors, 7.22, 7.24, 7.36
- Perron-Frobenius eigenvalue, 7.22, 7.24, 7.36
- Perron-Frobenius theorem, 7.22, 7.24, 7.36
- Perron-Frobenius theorem for nonnegative matrices, 7.29
- PGD, 8.2, 8.25
- PGM, 9.21
- photometric stereo, 0.20
- pilot signals, 2.27
- pivoting, 2.2
- POCS, 9.8
- POGM, 9.22
- polar decomposition, 5.36
- polar factorization, 5.36
- polylog, 9.5
- polynomial regression, 4.7
- poorly conditioned, 4.36
- positive, 7.35, 7.41
- positive (semi)definite matrices, 8.1
- positive definite, 2.35, 2.37, 5.8, 7.20, 8.3, 8.6, 8.13, 8.24, 8.27
- positive matrix, 7.20
- positive semi-definite, 2.35, 2.37, 7.20
- positive semidefinite, 8.3, 8.13, 8.27
- positive-definite, 2.2
- power iteration, 1.12, 7.17, 7.24
- power iteration convergence analysis, 7.19
- power method, 7.23
- preconditioned gradient descent, 8.2
- preconditioned steepest descent, 8.24
- preconditioner, 8.6
- preconditioning matrix, 8.2
- predict, 4.4
- primitive, 7.28, 7.32, 7.34, 7.36
- primitive matrix, 7.20, 7.24, 7.33

- principal component analysis, 6.21, 6.54
- principal square root, 8.3
- principle eigenvector, 1.12
- Procrustes problem, 5.2, 5.33, 5.40
- product, 1.67, 1.68
- projecting, 4.53
- projection, 4.52, 4.53, 8.30, 9.8
- projection matrix, 4.49, 7.33
- projections onto convex sets, 9.8
- proximal gradient method, 9.21
- proximity operation, 9.21
- pseudo-inverse, 4.23, 4.32
- push-through identity, 1.44
- Pythagorean theorem, 1.47

- QR decomposition, 2.2, 4.25
- quadratic, 8.2

- random walk, 7.40
- range, 3.21, 3.29, 4.53
- rank, 2.19, 2.39, 3.3, 3.23, 3.25, 3.51, 6.22
- rank 1 matrix, 1.22

- rank constraint, 9.12
- rank regularizer, 9.12
- rank-1 approximation, 6.14, 6.15
- Rayleigh quotient, 6.54
- rectangular diagonal matrix, 1.15, 2.18, 4.21
- recurrent neural networks, 5.27
- reducible, 7.34
- regularization parameter, 4.41
- regularizer, 6.35
- REPL, 0.12, 0.13
- residual, 4.6, 4.28, 4.29
- reverse triangle inequality, 5.6
- ridge regression, 4.37, 4.41
- right eigenvector, 1.9
- right eigenvectors, 2.38
- right inverse, 1.43, 4.19, 5.31
- right singular vector, 2.28, 2.34
- right singular vectors, 2.18
- risk, 6.45
- roots, 7.5
- rotation, 2.14

- rotation matrix, 2.8, 2.21
- row rank, 3.22, 3.23, 4.44
- row space, 3.21

- sampling mask, 9.6
- scaling, 5.8
- scatter matrix, 1.59
- Schatten 2-norm, 5.14
- Schatten p-norm, 5.20, 5.22, 5.25
- Schur complement, 1.44
- Schur decomposition, 2.2
- Schur norm, 5.14
- Schwarz, 1.48, 5.10
- scores, 6.55
- scree plot, 6.7, 6.10
- sensor localization, 6.23
- Sherman-Morrison-Woodbury identity, 1.44
- Sherman-Morrison formula, 1.44
- shift invariance, 1.16
- shrinkage, 6.35
- signal to noise ratio, 2.26
- similar, 2.9

- simple, 7.36
- simultaneously diagonalizable, 2.10
- simultaneously triangularizable, 8.21
- singular, 2.3
- singular matrix, 1.59
- singular value, 2.17
- singular value decomposition, 2.9, 2.18
- singular value hard thresholding, 6.37, 9.18
- singular value soft thresholding, 6.38, 9.19
- singular values, 2.17, 2.18, 4.44
- singular vectors, 2.17, 6.20
- smooth convex, 8.25
- SNR, 2.26
- span, 3.7–3.9, 3.12
- spark, 3.51
- sparse matrix, 1.16
- sparse subspace clustering, 6.59
- sparsity, 4.27
- spectral norm, 2.25, 5.18, 5.24, 5.25, 6.16, 8.26
- spectral radius, 5.26, 5.27, 5.45, 7.22
- spectral theorem, 2.4, 2.6

- square, 2.18, 7.22, 7.28
- square integrable, 1.69
- square matrix, 1.15
- SSC, 6.59
- stable rank, 6.22
- standard basis, 3.44
- states, 7.26
- stationary distribution, 7.29
- steady state, 7.30
- steady-state distribution, 7.29
- steepest descent, 8.24
- Stein's unbiased risk estimate, 6.45
- step size, 8.2, 8.29
- Stiefel manifold, 5.40, 5.44
- stochastic eigenvector, 7.29, 7.31
- strictly convex, 5.6, 8.27, 8.33
- strongly connected, 7.37, 7.38, 7.41
- strongly connected graph, 7.39
- sub-multiplicative, 5.13–5.17, 5.22, 5.25, 5.45
- subadditive, 3.27
- subspace, 3.3, 3.4, 3.6, 3.7, 3.19, 3.29
- subspace clustering, 6.59
- subspace learning, 6.59
- subspace sum, 3.16
- subspace union, 3.16
- sum, 1.67, 1.68, 3.15
- sum of outer products, 1.36
- supervised learning, 6.59
- supervised PCA, 6.58
- supremum, 5.3
- SURE, 6.45
- surrogate function, 9.14
- SVD, 1.15, 2.2, 2.3, 2.9, 2.18, 2.39, 2.40, 4.21–4.23, 4.42, 4.60, 5.31, 5.33, 5.35, 5.42, 5.45, 6.2, 6.20, 6.53, 6.54, 9.17
- SVST, 6.38
- Sylvester's determinant identity, 1.56
- Sylvester's rank inequality, 3.27
- symmetric, 1.19, 3.5
- system of linear equations, 1.5, 4.2
- tall, 2.18, 2.23, 3.34, 3.35, 3.38, 3.48
- Taylor's theorem, 8.26

- term, 0.8
- term-document matrix, 1.7
- thin SVD, 2.23
- thresholding, 6.35
- tight frame, 1.49, 4.45, 4.46, 4.48
- tight upper bound, 2.25
- Tikhonov regularization, 4.37, 4.41
- time invariance, 1.16
- time-homogeneous Markov chain, 7.26
- Toeplitz, 1.16
- Toeplitz matrix, 7.15
- trace, 1.65
- trace norm, 5.14, 5.20
- training, 6.52
- transient, 7.39
- transition, 7.46
- transition matrix, 7.27, 7.29, 7.32
- transition matrix., 7.27
- transition probabilities, 7.26
- translation, 5.40
- transpose, 1.19, 1.20, 1.69
- Triangle inequality, 1.47
- triangle inequality, 5.2, 5.6, 5.13, 8.26
- tridiagonal, 1.14
- trigonometric identities, 3.40
- truncated SVD, 4.35, 4.36, 4.41, 6.6, 6.32
- tune, 4.41
- twice differentiable, 8.26

- uncorrelated, 5.11
- under-determined, 4.26
- union, 3.16
- union of subspaces, 3.16
- unique, 6.13, 7.32, 9.9
- unit norm, 1.46
- unit-norm, 1.46
- unitarily invariant, 5.7, 5.25, 5.32, 5.45, 6.5, 6.16, 6.23, 6.35
- unitary, 1.50, 1.62, 2.4, 2.6, 3.28, 6.19
- unitary eigendecomposition, 2.6, 2.8, 4.50
- unitary matrix, 1.49, 3.45, 4.46, 4.47, 7.11
- unity, 1.67
- unsupervised, 6.59

upper Hessenberg, [1.15](#)
upper triangular, [1.14](#), [7.4](#)

Vandermonde matrix, [7.7](#)

vector, [1.3](#), [1.20](#), [1.66](#)

vector 2-norm, [5.3](#)

vector addition, [1.68](#)

vector norm, [5.2](#)

vector norms, [5.2](#), [5.29](#), [5.45](#)

vector space, [1.66](#), [1.68](#), [3.4](#), [3.20](#), [5.9](#), [5.13](#)

vectors, [1.68](#)

Venn diagram, [4.52](#)

weakly differentiable, [6.45](#)

weighted, [5.4](#)

weighted 2-norm, [5.4](#)

wide, [2.18](#), [2.23](#), [3.34](#), [3.36](#)

Wikipedia, [0.8](#)

zero, [1.67](#)

zero vector, [1.68](#)