

---

## Subject index

- $L^1[0, 1]$ , 384
- $M$ -estimator
  - multivariate regression, 266
- $U$ -statistics
  - tail bounds, 38
- $\ell^2(\mathbb{N})$ , 130, 384
- $\ell_1$ -regularization, 227
  - basis pursuit, 227
  - generalized linear models, 261
  - Ising model selection, 368
- $\ell_q$ -ball, 428
- $q$ -convex hull, 428
- $f$ -divergences
  - equivalent forms, 521
- $k$ -clique problem, 256
- additive matrix decomposition, 343
- approximation
  - error, 411, 433, 434
  - theory, 154
- Azuma–Hoeffding inequality, 36
- basic inequality
  - general  $M$ -estimator, 274
  - Lasso, 212
  - modified nonparametric, 445
  - nonparametric least squares, 423, 431
  - PCA, 243
  - square-root Lasso, 234
- basis pursuit, 227
- Bayes error, 491
- Bayes risk, 491
- Bennett’s inequality, 51
- Bernstein’s condition
  - random matrix, 171
  - random variable, 27
- Bernstein’s inequality, 28
  - matrix version, 176, 191, 192
  - uniform version, 89
  - vector version, 192
- Bochner’s theorem, 411
- bounded differences inequality, 37
- Brunn–Minkowski inequality, 72, 94
- Cauchy sequence, 383, 384
- chaining, 139, 154
  - Dudley’s entropy integral, 140
  - generic, 154
- Chebyshev’s inequality, 21, 50
- Chernoff bound, 22
- chi-squared variable, 29, 30, 43
- classification, 299
- classification
  - logistic regression, 262, 299, 471
  - support vector machine, 299, 415, 472
- CLIME estimator, 380
- clique, 347
  - compatibility function, 347
- compressed sensing, 198
- concentration
  - Brunn–Minkowski inequality, 72
  - concentration function
    - Lipschitz, 70
  - data compression, 55
  - Dvoretzky–Kiefer–Wolfowitz theorem, 117
  - function, 68
    - sphere, 68
  - Gaussian complexity, 43
  - Herbst argument, 60
  - Hilbert space, 54
  - Ising model, 56
  - kernel density estimation, 54
  - Lévy’s bound, 69
  - Lipschitz function, 70, 432
  - means and medians, 53
  - Orlicz norm, 150
  - Pinsker–Csiszár–Kullback inequality, 78
  - Prékopa–Leindler inequality, 74
  - Rademacher complexity, 40, 63
  - singular values, 64
  - spin glasses, 56
  - sub-exponential, 26
  - sub-Gaussian, 23
  - suprema of Orlicz process, 151
  - uniform laws, 106
- conditional independence, 350
- convex program
  - basic inequality, 213, 423, 431
  - constrained Lasso, 206
  - Lasso, 206
  - linear program, 200
  - perturbation theory, 210

- primal–dual witness, 258
- quadratic program, 206
- relaxed basis pursuit, 206
- SDP, 258
- semidefinite program, 258
- convex regression, 430, 458
- corrected regression, 370
- coupling, 77
- Courant–Fischer theorem, 256
- covariance matrix
  - $\ell_q$ -sparsity, 184
  - graph-structured, 182
  - sample, 160
  - sparse, 180
  - spiked model, 255
- covering number, 122
  - $\ell_1$ -ball, 149
  - binary hypercube, 123
  - Lipschitz function class, 127
  - parametric function class, 126
  - relation to packing number, 124, 154
  - relation to VC dimension, 155
  - relation to volume ratios, 125
  - singular values, 157
  - Sobolev class, 129
  - unit balls, 126
  - unit cube, 123
- Dantzig selector, 227
- data compression, 55
  - principal component analysis, 239
- decomposability, 270
  - $\ell_1$ -norm, 270
  - group norm, 271
  - key consequence, 273
  - nuclear norm, 318
  - overlapping group norm, 271
- density estimation, 484
- design matrix
  - irrepresentable condition, 229
  - pairwise incoherence, 227
  - restricted eigenvalue, 208
  - restricted eigenvalue condition, 213, 228
  - restricted nullspace, 227
  - restricted nullspace property, 202
- direct sum, 402
- discretization
  - basic bound, 135
  - Gaussian complexity of unit ball, 137
  - sub-Gaussian random matrix, 137, 166
- discriminant analysis, 3
- Dudley’s entropy integral, 140, 154
  - $\ell_q$ -“balls”, 429
  - Lipschitz function class, 430
  - nonparametric least squares, 426
  - Orlicz process, 151
  - parametric function class, 142
  - ridge regression, 427
  - sparse linear regression, 429
  - uniform law, 457
  - VC class, 142
- Dvoretzky–Kiefer–Wolfowitz theorem, 117
- Eckart–Young–Mirsky theorem, 237
- eigenvalues
  - Courant–Fischer representation, 256
  - perturbation, 241
  - spectral gap, 242
- empirical process theory
  - chaining, 140
  - Dudley’s entropy integral, 140
  - Glivenko–Cantelli theorem, 100, 111
  - localization, 448
  - peeling, 302, 336, 447, 448
  - polynomial discrimination, 110
  - Rademacher complexity, 105
  - symmetrization, 107
  - uniform laws, 98
  - Vapnik–Chervonenkis dimension, 111
  - VC class, 112
- entropic method
  - Bernstein’s inequality, 61
  - general  $\phi$ -entropy, 58
  - Herbst argument, 60
  - separately convex functions, 62
  - sub-Gaussian bound, 60
  - tensorization, 64
- entropy
  - bounded random variables, 93
  - constant shifts, 93
  - exponential families, 93, 94
  - rescaling, 93
  - Shannon, 92
    - Gaussian distributions, 522
  - variational representation, 92
- evaluation functional, 390
- exponential families, 365
  - concentration, 94
  - entropy, 94
- Fano’s inequality, 517
- Fano’s method, 500
  - density estimation, 504, 513
  - Gaussian location family, 502
  - linear regression, 503
  - local packing, 504
  - local packing approach, 503
  - principal component analysis, 509
  - sparse linear regression, 506
  - sparse PCA, 511
  - sparse variable selection, 507
  - with Gaussian entropy bounds, 506
  - Yang–Barron version, 512
- filtration, 33
- fixed design, 417
- Fourier basis, 435
- Fourier coefficient, 398

- free energy, 56
- function
  - Lipschitz, 70
  - matrix monotone, 169, 190
  - restricted strong convexity, 277
  - separately convex, 62
  - strongly convex, 277
- function class
  - $b$ -uniformly bounded, 454
  - Lipschitz, 127, 139, 429
  - parametric, 126, 142
  - Sobolev, 129
  - star hull, 424
  - star-shaped, 424, 454
- functional
  - ANOVA decomposition, 402
  - bounded linear, 385
  - continuous in sup-norm, 118
  - evaluation, 390
  - linear, 385
- gauge function, 256
- Gaussian
  - $\phi$ -entropy, 59
  - canonical process, 132
  - chaos variable, 44, 54
  - complexity, 43, 117, 132, 306
    - $\ell_0$ -“balls”, 133, 156
    - $\ell_1$ -norm ball, 133
    - $\ell_q$ -balls, 155
    - ellipsoid, 156
    - Euclidean ball, 132, 148
    - Lipschitz functions, 139
    - relation to Rademacher, 155
    - unit ball, 137
  - concentration for Lipschitz functions, 40, 432
  - concentration for suprema, 157
  - contraction inequality, 147, 157
  - covariance selection, 352
  - graphical Lasso, 353
  - graphical model, 265, 352
    - covariance selection, 352
  - kernel function, 398, 411
  - Kullback–Leibler divergence, 521
  - local complexity, 421
    - critical inequality, 422
  - location family, 492
  - maximum Shannon entropy, 522
  - Mills ratio, 50
  - mixture model, 240, 257, 366
  - moment generating function, 22
  - random matrix, 44, 157
  - sequence model, 196, 227
    - minimax risk, 227
  - upper bounds for maxima, 53
- Gaussian comparison inequality, 144, 228
  - Gordon, 154, 185, 186
  - random matrix, 157, 185
  - Slepian, 145, 154
  - Sudakov minoration, 148
  - Sudakov–Fernique, 146, 148, 154
- generalized linear models, 261, 278, 308
- Glivenko–Cantelli
  - class, 101
  - classical theorem, 100, 111
  - failure, 118, 119
  - relation to Rademacher complexity, 105
- graphical Lasso, 265, 306, 353
  - corrected version, 370, 381
- graphical model
  - exponential families, 365
  - factorization, 347–349
  - Gaussian, 265, 349, 352
  - Hammersley–Clifford theorem, 265, 351
  - Ising model, 56, 349, 380
  - Markov property, 350
  - mixed variables, 366
  - neighborhood cutsets, 350
  - parameter estimation, 352
  - Poisson variables, 366
  - Potts model, 365, 366
  - selection, 352
    - Gaussian case, 352
    - Ising case, 367
  - Sherrington–Kirkpatrick model, 56
  - undirected, 347
- Hammersley–Clifford theorem, 265, 351
- Hanson–Wright inequality, 54
- Hellinger distance, 490
  - and Kullback–Leibler, 484
- decoupling, 520
- Hilbert space, 384
  - $L^2[0, 1]$ , 384
  - $\ell^2(\mathbb{N})$ , 384
  - concentration, 54
  - nullspace of bounded linear functional, 412
  - projections, 412
  - separable, 384
- Hilbert–Schmidt operator, 394
- hinge cost, 299, 415, 472
- histogram estimator, 477, 484
- Hoeffding’s inequality, 24
  - matrix version, 174
  - uniform version, 87
- inner product, 383
- integral probability metric, 412
- irrepresentable condition, 234
  - graphical Lasso, 357
- Ising model, 56, 349
  - conditional distributions, 380
- graph selection, 367
- isoperimetry
  - discrete, 72
- Johnson–Lindenstrauss embedding, 30, 49
- Kantorovich–Rubinstein duality, 77, 92

- kernel
  - Cauchy–Schwarz inequality, 413
  - density estimation, 54
  - discrete space, 395
  - Gaussian, 386, 398, 415, 422, 442, 459
  - linear, 386
  - Mercer’s theorem, 396
  - polynomial, 386, 413, 440
  - positive semidefinite, 386
  - power set, 413
  - probability space, 413
  - Sobolev, 422, 440
  - Sobolev space, 395, 397
  - uniqueness, 389, 412
  - universal, 412
- Kullback–Leibler divergence, 52, 92, 490
  - and Hellinger distance, 484
  - and Shannon entropy, 92
  - chain rule, 92
  - mixture representation, 521
  - multivariate Gaussian, 521
  - properties of, 519
- Löwner–Heinz theorem, 169, 190
- Lasso, 361
  - $\ell_2$ -bounds, 233
  - $\ell_\infty$ -bounds, 232, 233
  - constrained form, 418
  - corrected version, 371, 381
    - additive corruptions, 381
    - missing data, 382
  - generalized linear, 262
  - graphical, 265, 353
    - corrected version, 370
    - edge selection, 357
    - Frobenius bounds, 355
    - operator norm bounds, 357
  - group, 262
  - irrepresentable condition, 234
  - multivariate group, 266
  - non-convex version, 372
  - overlapping group, 271, 293, 307
  - pairwise incoherence, 234
  - population cost, 260
  - primal–dual witness, 229
  - special case of  $M$ -estimator, 260
  - square-root, 229, 234
  - variable selection, 229
  - weighted version, 233
- Le Cam’s inequality, 490, 520
- Le Cam’s method
  - Gaussian location family, 492, 498
  - pointwise density estimation, 494
  - quadratic functionals, 496, 499
  - uniform location family, 493
- Lieb’s inequality, 173
- linear functional
  - bounded, 412
- Lipschitz function, 40, 429
  - concentration for Gaussian variables, 40
  - contraction inequality, 147, 157
- Lipschitz functions, 484
- $M$ -estimators, 228
- Marčenko–Pastur law, 7
- Markov properties
  - effect of additive noise, 380
  - neighborhood regression, 380
- Markov’s inequality, 21, 50
- martingale, 33
  - difference sequence, 35
  - Doob construction, 34
  - likelihood ratio, 34
- mass transportation, 92
- matrix
  - $\ell_2$ -operator norm, 6, 44, 63
    - analysis, 255
    - approximation, 237
      - low-rank, 238
  - completion problem, 177, 313
  - compressed sensing, 315
  - decomposition problem, 267
  - eigengap, 242
  - Frobenius norm, 44, 312
  - Gaussian random, 44
  - inverse covariance, 349, 380
  - monotone function, 169, 190
  - non-negative and operator norms, 190
  - nuclear norm, 268, 313
  - operator norm, 6, 63, 190
  - perturbation theory, 255
  - power method, 257
  - precision, 265, 349, 380
  - regression, 312
  - sample covariance, 160
  - singular values, 157
  - spectral norm, 6, 44, 63
  - spiked covariance, 245
    - sparse version, 252
  - Toeplitz, 238
  - trace inner product, 312
  - tri-diagonal, 352
  - unitarily invariant norm, 256
- maximum likelihood
  - uniform laws, 118
- Mercer’s theorem, 395, 396
  - discrete space, 396
  - polynomial kernel, 396
  - Sobolev, 397
- metric
  - Euclidean, 121
  - Hamming, 121
  - Hellinger, 484
  - integral probability, 409, 412
  - Kolmogorov distance, 409
  - total variation, 410

- Wasserstein, 76
- metric entropy, 123
  - $\ell_1$ -ball, 149, 154
  - $\ell_q$ -ball, 428, 449
  - $q$ -convex hull, 428, 449
  - chaining, 140
  - Dudley's entropy integral, 140
  - ellipsoid, 130, 154
  - Lipschitz function class, 127
  - parametric function class, 126
  - relation to VC dimension, 155
  - relation to volume ratios, 125
  - RKHS, 130
  - Sobolev class, 129, 130
  - unit balls, 126
  - VC class, 154
- metric space, 121
  - $C[0, 1]$ , 122
  - $L^2(\mu, [0, 1])$ , 122
  - complete, 384
  - covering number, 122, 154
  - measure, 67
  - packing number, 124, 154
  - totally bounded, 154
- Mills ratio, 50
- minimax risk, 486
  - additive nonparametric regression, 523
  - computational constraints, 519
  - constrained versions, 519
  - density estimation, 504, 513
  - Fano's method, 500, 502
  - Gaussian location family, 502
  - Gaussian sequence model, 227
  - generalized linear models, 522
  - Le Cam's method, 498
  - linear regression, 503
  - pointwise density estimation, 494
  - principal component analysis, 509
  - private versions, 519
  - quadratic functionals, 496, 499
  - sparse linear regression, 506
  - sparse PCA, 255, 511, 522
  - sparse variable selection, 507
- Minkowski sum, 72
- mixture model
  - Gaussian components, 257
- neighborhood selection, 361
- non-convex optimization, 227, 372
- nonparametric
  - density estimation, 54, 475, 513
  - least squares, 416
  - maximum likelihood, 484
  - regression, 7, 417
    - basic inequality, 431
  - variable selection, 229
- norm
  - $L^2(\mathbb{P})$ , 453
  - $L^2(\mathbb{P}_n)$ , 453
  - $L^2(\mu, [0, 1])$ , 122
  - $\ell_1$ , 231
  - atomic, 306
  - decomposable, 270, 317
    - $\ell_1$ -norm, 270
    - group norm, 271
    - nuclear norm, 318
  - dual, 272, 306, 308
  - empirical  $L^2$ , 453
  - Euclidean, 121
  - Frobenius, 312
  - group sparse, 262
  - Hamming, 121
  - nuclear, 268, 313
  - operator, 6, 63, 190
  - overlapping group, 294, 307
  - overlapping group sparse, 264
  - sup-norm, 122
  - total variation, 76, 410, 414, 519
  - Wasserstein distance, 95
  - weighted  $\ell_1$ , 230, 233
- optimal transport, 92
- oracle inequality, 280
  - Fourier expansions, 436
  - general  $M$ -estimator, 280
  - matrix decomposition, 340
  - nonparametric least squares, 433
  - nuclear norm regularization, 320
  - orthogonal series expansion, 434
  - sparse regression, 436
- order statistic, 43
- Orlicz
  - concentration, 150
  - Dudley's entropy integral, 151
  - maxima of random variables, 55
  - norm, 55, 150
  - process, 151
  - random variables, 55
- orthogonal series, 434
- packing number, 124
  - Boolean hypercube, 155
  - relation to covering number, 124, 154
  - unit cubes, 124
- pairwise incoherence, 227, 230, 231, 234
- Parseval's theorem, 384
- peeling technique, 302, 336, 447, 448
- perturbation
  - bound for eigenvalues, 241
  - bound for eigenvectors, 243
  - eigenvalues, 241
  - eigenvectors, 241
  - role of eigengap, 242
  - theory for operators, 255
- phase retrieval, 257, 315, 327, 343
- Pinsker–Csiszár–Kullback inequality, 79, 490, 520
- plug-in principle, 3

- corrected Lasso, 371
- principal component analysis, 237
- population cost function, 259
- Potts model, 365
- Prékopa–Leindler inequality, 74
- precision matrix, 265
  - CLIME estimator, 380
  - graphical Lasso, 358
- primal–dual witness, 229
  - Gaussian graph selection, 229
  - graphical Lasso, 229, 358
  - group Lasso, 229
  - Lasso, 229
  - Lasso with random design, 362
  - logistic regression, 368
  - nonparametric variable selection, 229
  - sparse PCA, 258
- principal component analysis, 236
  - basic inequality, 243
  - eigenfaces, 239
  - Gaussian mixture model, 240
  - lower bounds, 509
  - mixture models, 257
  - phase retrieval, 257
  - plug-in principle, 237
  - sparse eigenfaces, 248
  - spiked covariance, 245
    - with sparsity, 252
  - with sparsity, 248, 258
- Rademacher
  - chaos variable, 95
  - complexity, 40, 105, 117, 132
    - $\ell_1$ -norm ball, 133
    - concentration, 63
    - Euclidean ball, 132
    - kernel classes, 448
    - local empirical, 448, 455, 480
    - local population, 448, 454, 480
    - polynomial discrimination, 110
    - properties, 119
    - relation to Gaussian, 155
    - VC-based upper bounds, 109
  - contraction inequality, 147
  - desymmetrization, 107
  - random variable, 23
  - symmetrization, 107
- random design, 229, 417
- random matrix
  - Bernstein’s condition, 171, 191
  - Gaussian comparison inequality, 158, 164, 185, 188
  - Gaussian ensemble, 44, 161
  - singular values, 44, 63, 157
  - sub-Gaussian, 174, 190
  - Sudakov minoration, 149
  - symmetrization, 173
  - variance of, 190
  - Wishart, 161
- random projection, 30, 198, 211
- random variable
  - $\chi^2$ , 42
  - chi-squared, 29, 30, 43
  - Gaussian, 22
  - Gaussian chaos, 54
  - Rademacher, 23
  - sub-exponential, 26, 48
  - sub-Gaussian, 23, 25, 45
- rate distortion, 55
- Rayleigh–Ritz formula, 160
- regression
  - additive nonparametric, 452, 463, 466, 484
  - approximation error, 433
  - constrained  $\ell_q$ , 418
  - convex, 419, 420, 449
  - convex Lipschitz, 430
  - corrected version of neighborhood-based, 370
  - estimation error, 433
  - Gaussian kernel, 442
  - isotonic, 449
  - kernel ridge, 407, 419, 439, 440
  - linear, 194, 427
  - matrix completion, 315
  - matrix version, 312
  - monotonic, 420, 449
  - multivariate, 265, 313
    - group, 267, 291
    - overlapping group, 267, 293
  - neighborhood-based, 360, 366, 380
  - noiseless linear, 194
  - oracle inequality, 433
  - orthogonal series, 434
  - phase retrieval, 315, 327
  - polynomial functions, 440
  - random design, 417
  - regularized estimator, 439
  - regularized nonparametric, 418
  - ridge, 418
  - shape-constrained, 449
  - smoothing spline, 419
  - Sobolev space, 442
  - sparse linear, 428, 436
    - $\ell_0$ -“balls”, 436
    - $\ell_q$ -“balls”, 232, 429
  - with corrupted covariates, 370
- regularizer
  - $M$ -estimator, 260
  - $\ell_0$ -based, 235
  - $\ell_1$ -norm, 261
  - decomposable, 270, 306
  - group norm, 262, 307
  - non-convex, 227
  - nuclear norm, 268
  - overlapping group norm, 264, 307
  - weakly decomposable, 306
- representer theorem, 408

- reproducing kernel Hilbert space, 390
  - direct sum, 412
  - eigenfunctions, 396
  - ellipsoid, 400
  - Fourier expansion, 387
  - function interpolation, 405
  - Gaussian kernel, 386, 398, 442
    - regression, 442
    - uniform law, 459
  - kernel integral operator, 394
  - kernel ridge regression, 419
  - kernel trick, 387, 388
  - linear functions, 391
  - linear kernel, 386, 413
  - Mercer's theorem, 396
  - operations on, 400
  - polynomial kernel, 386, 396, 413, 440
  - positive semidefinite kernel, 386
  - representer of evaluation, 390
  - representer theorem, 408, 411
  - smoothing spline, 395
  - smoothing splines, 393, 397
  - Sobolev space, 392, 393, 395, 397, 401, 442
    - uniform law, 459
  - translation-invariant kernels, 397
  - uniform law, 458
  - uniqueness of kernel, 412, 413
- restricted
  - curvature condition, 276, 284
    - dual norm form, 284
    - operator norm, 320
  - eigenvalue condition, 207, 208, 213, 228
    - $\ell_1$ -version, 230
    - $\ell_2$ -version, 208, 228, 230–232, 278
    - $\ell_\infty$ -version, 233, 285
  - isometry property, 227, 230, 231
  - nullspace condition, 213
  - nullspace property, 202, 227
  - strong convexity, 277, 319
    - error bounds, 280
    - generalized linear models, 278
    - graphical Lasso, 355
    - group Lasso, 290
    - least-squares, 278
    - low-rank matrix, 319
    - overlapping group, 294
- ridge regression, 418
- Riesz representation theorem, 385
- RKHS, 419
  - feature map, 396
  - Gaussian kernel, 415
  - kernel ridge regression, 440
  - kernel trick, 411
  - polynomial kernel, 413
  - universal, 412
- Rosenthal's inequality, 55
- sample covariance, 5
- semidefinite program, 255
- Shannon entropy, 92
  - chain rule, 520
  - properties of, 519
- Shannon's rate distortion theorem, 55
- shatter coefficient, 112
- Sherrington–Kirkpatrick model, 56
- Slepian's inequality, 145
- smoothing spline, 419
- Sobolev space
  - regression, 442
  - reproducing kernel Hilbert space, 392
- sparse PCA, 248, 255, 258
  - hardness result, 256
  - lower bounds for  $\ell_2$ -estimation, 522
  - lower bounds for variable selection, 511, 522
- sparsity
  - $\ell_0$ -“balls”, 235
  - $\ell_q$ -“balls”, 195, 229, 418
    - covariance matrices, 184
    - vector, 195
  - compressed sensing, 198, 211
  - graph structure, 182
  - hard, 195
    - covariance matrices, 181
    - vector, 195
  - models, 194
  - weak, 195
- spiked covariance matrix, 255
- spin glass, 56
- spline, 419
- star-shaped, 424
- sub-exponential, 26
  - Bernstein's condition, 27
  - Bernstein's inequality, 28
  - chi-squared, 29, 30, 43
  - equivalent characterizations, 31, 48
  - Johnson–Lindenstrauss embedding, 30
  - operations, 29
  - tail bound, 26
- sub-Gaussian, 23, 51
  - bounded random variables, 51
  - equivalent characterizations, 25, 45
  - lower tail bounds, 51
  - operations on variables, 53
  - random matrix, 174
    - discretization argument, 166
  - random variable, 23
  - upper bounds for maxima, 53
- Sudakov minoration, 148, 149
  - singular value, 149
- Sudakov–Fernique inequality, 146, 157, 164
- support vector machine, 299, 411, 415, 472
- symmetrization, 24, 107
  - random matrix, 173
- tail bounds
  - $U$ -statistics, 38

- Azuma–Hoeffding inequality, 36
- Bennett’s inequality, 51
- Bernstein’s inequality, 28, 52
  - matrix version, 176
  - uniform version, 89, 96
- bounded differences inequality, 37
- Chebyshev’s inequality, 21, 50
- Chernoff inequality, 22
- Dvoretzky–Kiefer–Wolfowitz theorem, 117
- for order statistics, 43
- Gaussian concentration for Lipschitz functions, 40
- Hanson–Wright inequality, 54
- Hoeffding’s inequality, 24
  - matrix version, 174
  - uniform version, 87
- lower binomial bounds, 52
- lower bounds for non-negative RVs, 31
- lower for squared sub-Gaussians, 51
- Markov’s inequality, 21, 50
- martingale difference sequences, 35
- matrix completion, 177
- Mills ratio, 50
- moment-based, 51, 55
- non-symmetric random matrices, 191
- one-sided Bernstein’s inequality, 31
- Rosenthal’s inequality, 55
- sub-exponential, 26
- upper binomial bounds, 52
- tensor product, 403
- thresholding
  - hard, 180, 197, 229
  - soft, 197, 229, 310
  - sparse PCA, 255
- total variation distance, 489, 519
  - mixture distributions, 521
  - Wasserstein distance, 95
- truncation argument, 468
- uniform law
  - $L^2(\mathbb{P})$ - and  $L^2(\mathbb{P}_n)$ -norms
    - one-sided, 464
    - two-sided, 455
  - convex Lipschitz functions, 457
  - Dudley’s entropy integral, 457
  - Gaussian kernel, 459
  - Glivenko–Cantelli theorem, 100, 111
  - kernel classes, 458, 482, 483
  - Lipschitz cost functions, 473
  - quadratic functions, 455
  - random matrix, 161
  - restricted eigenvalues, 484
  - Sobolev space, 459
  - uniformly bounded functions, 105
- Vapnik–Chervonenkis theory
  - operations on VC classes, 115
  - shatter coefficient, 111
  - VC class, 112
  - Dudley integral, 142
- VC dimension, 111, 112
  - closed convex sets, 120
  - half-intervals, 112, 120
  - half-spaces, 116
  - infinite, 120
  - linear functions, 116
  - monotone Boolean conjunctions, 120
  - polygons, 120
  - relation to metric entropy, 155
  - spheres, 116, 120
  - two-sided intervals, 112
- VC–Sauer–Shelah bound, 113
- vector spaces, 115
- variational representation, 93, 94, 160, 161, 307
  - Courant–Fischer, 256
  - eigenvectors, 257
- vertex cutset, 350
- volume ratios, 125
- Wasserstein distance, 76
- Weyl’s inequality, 45, 189, 241, 256