

- mes, G., 1981. "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Transactions on Automatic Control*, vol. AC-26, pp. 301-320.
- mes, G., and B.A. Francis, 1983. "Feedback, minimax, sensitivity, and optimal robustness," *IEEE Transactions on Automatic Control*, vol. AC-28, pp. 585-601.
- evi, A.J., R. Meir, and V. Majorov, 1998. "Error bounds for functional approximation and estimation using mixtures of experts," *IEEE Transactions on Information Theory*, vol. 44, pp. 1010-1025.
- ki, S., 1993. *A Vision of the Brain*, Oxford: Blackwell Scientific Publications.
- psper, D., and D.E. Rumelhart, 1990. "The neurobiological significance of the new learning models," in *Computational Neuroscience*, E.L. Schwartz, ed., pp. 192-200, Cambridge, MA: MIT Press.

Index

- Activation function, 11, 168
 - definition of, 11
 - nonmonotonic, 519, 726, 731
 - types of, 12-15, 168-169
- Activation potential, see induced local field
- AdaBoost, 360, see also Boosting,
 - error performance, 362
 - summary of algorithm, 362
- Adaptive filtering, 118-120,
 - adaptive process, 120
 - filtering process, 120
- Adaptive pattern classification, experiments on, 187, 305, 337, 468
- Adaptive principal components extraction (APEX) algorithm, 422-429
- Adaptive resonance theory, 41, 795.
- Additive model, 650-651, 676-677
- Algorithm, origin of the term, 106
- An information-theoretic criterion, 253
- Anti-Hebbian learning, 56, 423
- Architectural graph, 18
- Artificial intelligence, 34
- Array antenna processing, ICA for, 513
- Associative Gaussian mixture model, 366
 - mixture of experts (ME) model, 368
 - probabilistic generative model, 367
- Asymptotic stability theorem, 406
- Attentional neurocomputing, 74, 793, 795
- Attractors, 674
 - basic of attraction, 675
 - fixed-point, 675
 - hyperbolic, 675, 774
 - manipulation of, 680
 - strange, 709-722
- Automata, finite-state, 747
- Autoregressive model, 31, 471
- Axon, 7
- Backgammon, 630
- Backprop, see Back-propagation algorithm
- Back-propagation algorithm, 161-175
 - accelerated convergence, 233-234
 - batch mode, 172
 - computational efficiency, 229-230
 - convergence, 231
 - delta rule, 166
 - generalized delta rule, 170
 - heuristics, 178-184
 - initialization, 182
 - learning rate, 169-171
 - local gradient, 163
 - local minima, 231
 - momentum, 170
 - output representation and decision rule, 184-187
 - scaling, 232
 - sensitivity, 230
 - sequential mode, 171-172
 - stopping criteria, 173
 - summary, 173-175
 - temporal, 652-658
 - virtues and limitations, 226-233
- Back-propagation through time, 751-756
 - computational complexity, 771
 - epochwise, 752-754
 - ordered derivative, 755-756
 - practical considerations, 755-756
 - truncated, 754-755
- Barlow's hypothesis, 504
- Bayes classifier, 143-148
 - Bayes risk, 143
- Bernoulli variables, 581
- Bias/variance dilemma, 87
 - approximation error, 88
 - estimation error, 88
- Biomedical records, ICA for, 513
- Bits, 486
- Blind deconvolution, 534
- Blind signal (source) separation, 72, 512
- Boltzmann distribution, see Gibbs distribution
- Boltzmann machine, 562-569
 - deterministic, 578-579
 - learning rule for, 60-61, 566-568
- Boosting, 357, 387
 - AdaBoost, 360
 - filtering method, 357
 - reweighting method, 357
 - subsampling method, 357
- Bounded, one-sided saturated function, 749

- Brain, 6
 - structural organization of levels, 9
- Brain-state-in-a-box (BSB) model, 703–709
 - clustering, 707–709
 - dynamics of, 706–707
 - Lyapunov function of, 705–706
 - network of networks using, 722
- Broyden-Fletcher-Goldfarb-Shanno algorithm, 244
- Cascade-correlation learning, 250
- Cauchy-Schwarz inequality, 140
- Cerebral cortex, cytoarchitectural map, 10
- Chaos, 709–722
 - correlation dimension, 713
 - definition of, 714
 - Lyapunov exponents, 713–714
 - dynamic reconstruction of, 174–718
- Chernoff bound, 193
- Church-Turing hypothesis, 748
- Classification and regression tree (CART), 374
- Cocktail party phenomenon, 72, 109, 534
- Cohen-Grossberg theorem, 701–703, 705
- Combinatorial optimization, 560
 - analogy with statistical physics, 561
- Committee machines, 351
- Competitive learning, 56, 294, 448
 - rule for, 59
- Computational complexity, 104, 292
 - exponential time algorithms, 347
 - polynomial time algorithms, 347
- Condition number, 132
- Conjugate-direction method, 238
- Conjugate-gradient method, 236–242
 - Brent's method, 242
 - comparison with Quasi-Newton's method, 244–245
 - Fletcher-Reeves formula, 239
 - line search, 240–242
 - Polak-Ribière formula, 239
 - residual, 239
 - summary of, 243
- Correctionism, 226–227
- Content-addressable memory, see Hopfield model
- Contextual maps, 474
- Continuous learning, 83, 750
- Convergence in probability, 91
- Convolution networks, 29, 245–247
- Correlation coefficient, 473, 507
- Correlation matrix, 127, 397
- Correlation matrix memory, 79–83
 - relation to LMS algorithm, 153
- Cortical (computational) maps, 9, 444, 477
- Cover's theorem on the separability of patterns, 257–261
- Cross-correlation vector, 128
- Cross-validation, 213–218
 - early stopping method, 215–217
 - generalized, 288
 - leave-one-out method, 218
 - model selection, 214–215
 - multifold cross-validation, 217–218
- Credit-assignment problem, 62, 164, 603
- Cumulants, 516
- Curse of dimensionality, 211–212, 291–292, 617
- Darwinian selective learning, 106
- Darmois' theorem, 543
- Davidon-Fletcher-Powell algorithm, 244
- Delta-bar-delta learning, 251, 253
- Dendrites, 7
- Deterministic annealing, 586–592
 - analogy with EM algorithm, 592
 - clustering, 586–591
 - Hidden Markov models, 596
 - pattern classification, 596
 - regression, 596
 - vector quantization, 596
- Differential entropy, 488
- Differentiation with respect to a vector, 150–151
- Dimensionally reduction, 401
- Dot product, see Inner product
- Dynamic programming, 603
 - asynchronous, 631
 - Bellman's optimality equation, 609–610
 - dynamic programming algorithm, 608–609
 - Gauss-Seidel method, 631
 - principle of optimality, 607–608
- Dynamic reconstruction 714–718
 - embedding delay, 715
 - method of false nearest neighbors, 716
 - recursive prediction, 716–717
 - Takens' theorem, 715
- Dynamical systems, 666–669
 - definition of, 666
 - Lipshitz condition, 668–669
 - state (phase) portrait, 667
 - state space, 666–668
- ϵ -insensitive loss function, 339–340
- Echo-locating bat, 1, 33
- Edgeworth expansion, 540
- Eigenvectors, 398
 - dominant, 403
- Empirical risk functional, 91
 - strict consistency, 92
- Empirical risk minimization, principle of, 92
- Ensemble averaging method of learning, 353, 387
- Entropy, in information-theoretic sense, 487
- Entropy, in thermodynamics sense, 548
- Equivariant property, 520–521
- Error back-propagation algorithm, see Back-propagation
- Error-correction learning, 51
- Error energy, 52
- Error-performance surface, 63
- Euclidean distance, 26
- Euler-Lagrange equation, 270–271
- Excitatory-inhibitory network, see Gradient descent-gradient ascent dynamics
- Expectation-maximization (EM) algorithm, 381–382, applied to HME model, 383
- Factorial distribution, 496, 581
- Feature space, 199, 258, 329
- Feedback, 14, 18
 - global, 664
 - local, 664, 786
- Feedforward networks, 21, 156, 256
 - fully-connected, 22
 - multilayer, 21
 - partially-connected, 22
 - single-layer, 21
- Financial market data analysis, ICA for, 513

- Finite-duration impulse response filters, 648
- Fisher's information matrix, 388
- Fisher's linear discriminant, 201–202
- Fletcher-Reeves formula, 239
- Fréchet differential, 268–270
- Free energy, 547
- Fuzzy systems, 793
- Gamma memory, 639–640
- Gauss-Newton method, see Optimization techniques, unconstrained
- Generalization, 2, 25, 205–208
 - training set size for, 208
- Generalized cross-validation, 287–289
- Generalized Hebbian algorithm (GHA), 414
 - convergence, 416
 - optimality of, 417
 - summary, 418
- Generalized Lloyd algorithm, 456
- Generalized sidelobe canceler, 74
- Gibbs distribution, 547, 594, 599
- Gibbs sampling, 561–562
 - convergence theorem, 562
 - ergodic theorem, 562
 - rate of convergence theorem, 562
- Global minima, definition, 249
- Gradient descent-gradient ascent dynamics, 724
- Gram-Charlier expansion, 515, 537–540
- Green's function, 271
- Green's identity, 270
- Green's matrix, 274
- Growth function, 94
- H₁ criterion, 151, 230
- Heaviside function, see Threshold function
- Hebbian learning, 55
 - covariance hypothesis, 57
 - generalized, 79
 - Hebb's postulate, 57, 394
 - synaptic enhancement, 56
- Hebbian synapse, 55
 - anti-Hebbian, 56
 - properties of, 55
- Helmholtz machine, 574–575
- Hessian matrix, 124, 204
 - computation of inverse, 224–225
- Hesteness-Stiefel formula, 254
- Hidden Markov models, 596, 643
- Hidden neuron, 21, 157
- Hierarchical clustering, 438
- Hierarchical mixture of experts (HME) model, 372
 - learning strategies for, 380
- Hierarchical vector quantization, 470
- Hilbert space, 269, 309
- Hopfield model (network), 680–696
 - energy function, 682
 - energy landscape, 686
 - fundamental memories (prototype states), 687
 - mixture states, 701
 - learning rule for, 690
 - load parameter, 694
 - retrieval (recall) phase, 689–690
 - reverse fundamental memories, 699–701
 - signal-to-noise ratio, 694
 - spin-glass states, 701
 - spurious states, 692–693
 - storage capacity, 693–696
- storage phase of learning, 688–689
- Hotelling's deflation technique, 416
- Hybrid systems, 37, 793
- Hyperbolic tangent function as activation function, 13, 169
- Identity map, see Replicator
- Image coding, 419
- Independent components analysis, 510–525
 - activation function for, 517–519
 - convergence considerations of learning algorithm, equivariant property, 520–521
 - learning algorithm for, 519–520
 - natural gradient for, 521
 - performance index for, 525
 - stability of learning algorithm, 521–522
- Induced local field, definition, 11
- Infimum, 91
- Influence matrix, 286
- Information preservation rule, 373
- Information-theoretic models of neural networks, 41
- Informon, 537
- Inner product, 26
- Inner-product kernel, 330, 433
- Inner-product space, 310
- Integrate-and-fire neuron, 725–726
- Intelligent machines, 790–794
 - for control, 792–793
 - for pattern recognition, 791–792
 - for signal processing, 793–794
- Interpolation theorem, 262–264
 - interpolation matrix, 264
- Inverse problems, 265
 - conditions for well-posedness, 266
- Iteratively reweighted least-squares, 389
- Jacobian matrix, 125, 204, 670
 - computation of, 202–204
- Jensen's inequality, 391
- Kalmar filters, 151, 762–765
 - conversion factor, 765
 - divergence phenomenon, 765
 - error-covariance matrix, 764
 - filtered estimation error, 765
 - innovations, 763
 - square root, 763
 - summary, 764
- Kalman filter, decoupled extended, 765–770
 - artificial process noise, 769
 - computational complexity, 770–771
 - multistream, 788
 - summary, 769–770
- Karhunen-Loève transform, see Principal components analysis
- Kernel matrix, 433
- Kernel principal components analysis, 432
 - summary, 435
- Knowledge, definition, 23
- Kullback-Leibler divergence (distance), 487, 495–497
 - pythagorean decomposition, 497
 - relation to mutual information, 496
- ~~Kushner-Clark~~ 407
- Lateral inhibition, 59
- Learning, 25
 - definition, 50
 - statistical theory, 84

- Learning tasks, 66
 - beamforming, 73
 - control, 70
 - filtering, 71
 - function approximation, 68
 - pattern association, 66
 - pattern recognition, 67
- Learning vector quantization, 467
- Learning with a teacher, 63
- Learning without a teacher, 64
- Least-mean-square (LMS) algorithm, 128–135
 - convergence, 130–132
 - LMS-Newton algorithm, 153
 - learning curve, 133–134
 - learning-rate annealing, 134–135
 - misadjustment, 133
 - normalized LMS algorithm, 152
- Least-squares filter, linear, 126–128
- Likelihood ratio, 145, 188
 - log-likelihood ratio, 146
- Likelihood ratio test, 145
- Line search, 240–242
- Linear separability, 138
- Linsker's model of mammalian visual system, 395
- Little model, 726
- Local minima, definition, 249
- Logistic function, 14, 45, 168
- Long-term potentiation (LTP), 107
- Lyapunov's theorems, 673–674
- Lyapunov function, 674
- Mahalanobis distance, 27
- Marginal entropy, 497
- Markov blanket, 583
- Markov chains, 548–556
 - Chapman-Kolmogorov identity, 550
 - classification, 555
 - definition, 548
 - ergodic, 551
 - ergodicity theorem, 552
 - irreducible, 550–551
 - principle of detailed balance, 555–556
 - recurrent properties, 550
 - state-transition diagram, 553
 - stochastic matrix, 549
 - transition probabilities, 549
- Markovian decision processes, 604–606
- Matrix inversion lemma, 225
- Maximum a posteriori (MAP) estimation, 389
- Maximum eigonfilter, Hebbian based, 404
- stability, 408
- Maximum entropy method for blind source separation, 529–533
 - equivalence with maximum likelihood, 531
 - learning algorithm, 532–533
- Maximum entropy (Max Ent) principle, 490
- Maximum likelihood estimation, 378
 - log-likelihood function, 379
 - properties, 388
- Maximum likelihood estimation for blind source separation, 525–528
 - relationship with independent components analysis, 527–528
- Maximum mutual information (Informex) principle, 484, 499–503
 - model for perceptual system, 504–505
 - relation to redundancy reduction, 503–505
- McCulloch-Pitts model, 14, 38, 135
- Mean-field theory, 576–578
- Memory, 75
 - associative, 67
 - correlation matrix, 79–83
 - crosstalk, 81
 - distributed, 75
 - long-term, 75
 - recall, 80
 - short-term, 75
- Memory, short-term structures, 636–640
 - memory depth, 638
 - memory resolution, 638
- Memory-based learning, 53
 - k-nearest neighbor rule, 54
 - nearest neighbor rule, 54
- Mercer's theorem, 331
- Method of Lagrange multipliers, 223, 323, 490
 - dual problem, 323, 328, 342
 - duality theorem, 324
 - Kuhn-Tucker conditions, 323
 - primal problem, 323, 328, 342
- Method of steepest descent, see Optimization techniques, unconstrained
- Metropolis algorithm, 556–558
- Michelli's theorem, 264–265
- Minimum description length (MDL) criterion, 253
- Minimum-norm solution, see Pseudoinverse
- Minor components analysis (MCA), 440
- Mixture of experts (ME) model, 368
- Model-reference adaptive control, 780–782
- Modularity, definition, 352
- Monomials, 259
- Multilayer perceptrons, 156
 - bounds on approximation error, 209–211
 - feature detection, 199, 227
 - feature space, 199
 - recurrent, 736–737
- Multinomial probability, 369
- Multivariate Gaussian functions (distributions), 275, 297, 492
- Mutual information, 492
 - for self-organized learning, 498
 - properties, 493
- NP-complete problems, 347
- Nadaraya-Watson regression estimator, 296, 479
- Natural gradient, 521, 540
- Nats, 486
- Necognitron, 108, 251, 795
- NETalk, 642–641
- Network pruning techniques, 218–226
 - approximate smoother, 221–222
 - complexity regularization, 219–222
 - optimal brain damage, 222
 - optimal brain surgeon, 222–226
 - weight decay, 220
 - weight elimination, 220
- Neural networks,
 - adaptivity, 3
 - architectures, 21
 - definition, 2, 17
 - fault-tolerance, 4
 - input-output mapping, 3
 - invariances built into, 29
 - neurobiological analogy, 4
 - properties, 2

- Neurodynamic programming, 603–634
 - finite-horizon problems, 606
 - infinite-horizon problems, 606
 - policy, 106
 - relation to reinforcement learning, 603
- Neuron, 7
 - models of, 10, 15
- Neuronal filters,
 - distributed, 648
 - focused, 644
- Neuromorphic systems, 5
- Newton's method, 235
- Neyman-Pearson criterion, 28
- Nonlinear principal components analysis, 434, 440
- Normed space, 267, 309
- Occam's razor, 206, 363
- Optimal brain surgeon algorithm, 226
- Optimal hyperplane, 320
 - quadratic method for computing, 322–325, 326
 - statistical properties, 325
- Optimization techniques, unconstrained, 121–126
 - Gauss-Newton method, 124–126
 - method of steepest descent, 121–122
 - Newton's method, 122–124
 - quasi-Newton methods, 242
- Ordered derivatives, 755
- Orthogonal similarity transformation, 399
- Outer product rule, see Hebbian learning
- Partition function, 547
- Perceptron, 135–143
 - relation to Bayes classifier, 143–148
- Perceptron convergence algorithm (theorem), 141
 - summary, 142
- Piecewise-linear function, 14, 703
- Plasticity, 1
- Polak-Ribière formula, 239
- Policy, 606
- Policy iteration, 610–612
 - approximate, 619–622
- Positive definite matrix, definition, 151
- Prediction, 72, 645, 771
- Principal components, definition, 400
- Principal components analysis, 396
 - adaptive methods, 431
 - batch methods, 431
 - decorrelating algorithms, 430
 - eigenstructure, 397
 - nonlinear, 434, 440
 - principal subspace, 430
 - reestimation algorithms, 430
- Principal curves (surfaces), 440, 461
- Principle of detailed balance, 555–556
- Principle of minimal free energy, 548
- Principle of minimum redundancy, 504
- Principle of orthogonality, 85, 402
- Principle of topographic map formation, 445
- Probably approximately correct (PAC) model, 102–105, 357
- Probability of correct classification, 191
- Probability of error (misclassification), 191
- Pruning, see Network pruning techniques
- Pseudo-differential operator, 276
- Pseudoinverse, 127, 284
- Pseudotemperature, 15, 547
- Q-factor, 610–611
- Q-learning, 622–627, 631–632
 - approximate, 624–625
 - convergence theorem, 623
 - exploration, 625–627
- Quadratic programming, 345
 - commercial libraries, 348
- Quasi-Newton method, 242
- Radial basis functions, 264
 - Gaussian, 264, 275, 297
 - inverse multiquadric, 264
 - multiquadric, 264
- Radial basis-function (RBF) networks, 256
 - approximation properties, 290–293
 - comparison with multilayer perceptron, 293
 - computational complexity, 292
 - generalized, 278–280
 - learning strategies, 298–305
 - normalized, 296
 - relation to kernel regression, 294
 - sample complexity, 292
- Random walk, 597
- Real-time recurrent learning, 756–762
 - computational complexity, 771
 - sensitivity graph, 761
 - summary, 760
 - teacher forcing, 762, 787
- Receptive fields, 28, 45, 87, 282
- Recurrent (neural) networks, 18, 23, 677–678
- Recurrent networks, dynamically driven, 732–789
 - computational power, 747–749
 - controllability and observability, 741–742
 - heuristics, 751
 - input-output model, 733–735
 - learning algorithms, 750–751
 - local controllability, 743–744
 - local feedback, 786
 - local observability, 744–746
 - network architectures, 733–739
 - nonlinear autoregressive with exogenous inputs, 746–747
 - recurrent multilayer perceptrons, 736–737
 - second-order models, 737–739
 - state-space model, 735–736, 739–746
 - vanishing gradients, 773–776
- Recursive least-squares (RLS) algorithm, 151
- Redundancy, 394, 503
 - measure for, 505
- Regression,
 - kernel, 294–298
 - nonlinear, 85, 285
 - ridge, 311
- Regression surface, 371
- Regularization network, 277–278
- Regularization theory, 219, 267
 - applied to dynamic reconstruction, 718
 - regularization parameter, 268, 284–290
- Reinforcement learning, 64–65, 603, 631
- Relative entropy, see Kullback-Leibler divergence
- Relative gradient, see Natural gradient
- Replicator, 227–229, 250–251
- Retina, 5
- Riemannian space, 540
- Riesz representation theorem, 269
- Robustness, 151, 230
- Rosenblatt's perceptron, see Perceptron

- Saddle point, 670
- Saliency, 223
- Sample complexity, 104
- Sauer's lemma, 99, 110
- Schlafl's theorem, 309
- Search-then-convergence learning schedule, 135
- Self-organization, 65, 393
 - principles of, 393
- Self-organizing maps (Kohonen's model), 446
 - batch version, 459
 - competitive process, 448, 478
 - conscience algorithm, 481
 - convergence phase, 453
 - cooperative process, 449
 - density matching, 460
 - neighborhood function, 450
 - ordering phase, 452
 - properties, 454
 - renormalized algorithm, 450, 483
 - summary, 453
 - synaptic adaptation, 451, 478
 - topological ordering, 459
- Semantic maps, see Contextual maps
- Sensitivity, 203, 230
- Shape-from-shading, 438
- Sigmoid belief networks, 569–574
 - deterministic, 579–586
 - learning rule, 571–573
 - mean-field distribution, 580
 - mean-field equation, 583
- Sigmoid function, 14
- Signal-flow graph, 15
 - basic rules, 16
- Singular value decomposition, 431
 - singular values, 431
 - singular vectors, 431
- Simulated annealing, 558–560
 - annealing schedule, 559–560
 - combinatorial optimization, 560–561
- Slack variables, 327, 341
- Smoother, 72
- Smoothness, measure of, 310
- Spatially coherent features, 506–508
- Spatially incoherent features, 508–510
- Spectral theorem, 399
- Spectrogram, 642
- Splines,
 - thin-plate, 312
- Stability, 672–673
 - Lyapunov's theorems, 673–674
- Stability-plasticity dilemma, 4
- Stagecoach problem, 614–617, 627–629
- State-space model of recurrent network, 739–746
- Statistical independence, 495
- Statistical mechanics, 546–548
- Stochastic machines rooted in statistical mechanics, 545–595
- Storage capacity of a surface, 261–262
- Stochastic approximation, 135
- Structural risk minimization, 100–102
- Sub-Gaussian distribution, 541
- Super-Gaussian distribution, 541
- Supervised learning, 63
 - as ill-posed hypersurface reconstruction problem, 265–266
 - as optimization problem, 234–245
- Support vectors, 321
- Support vector machines, 318
 - comparison with back-propagation learning, 338–339
 - optimum design, 332
 - pattern recognition, 329
 - regression, 340
- Subspace decomposition, 403
- Supremum, 91
- Synapse, 6
 - chemical synapse, 6
- Synaptic convergence, 16
- Synaptic divergence, 17
- System identification, 120, 659, 776–779
 - input-output model, 778–779
 - state-space model, 776–778
- Tapped-delay-line memory, 638–639
- TD-gammon, 631
- Temporal difference learning, 631
- Temporal processing, 635–663
 - network structures for, 640–643
- Threshold function, 12
- Tikhonov functional, 268
- Tikhonov-Phillips regularization, see Regularization theory
- Time, 635
 - explicit representation, 635
 - implicit representation, 635
- Time-delay neural network, 641–643
- Time-frequency analysis, 795
- Time-lagged feedforward networks, 636, 659
 - distributed, 651
 - focused, 643–646
 - universal myopic mapping theorem, 646–647
- Topographic maps, 8
- Travelling salesman problem, 597–598
 - solution using Hopfield model, 723–724
- Turing machine, 748
- Unit-delay operator, 19
- Universal approximation theorem, 208–209, 229
- Universal myopic mapping theorem, 646–647
- Unsupervised learning, 65
- Value iteration, 612–617
- Vanishing gradients problem, 773–776
- VC dimension, 94–98
 - bounds, 97, 110
 - definition, 95
- Vestibule-ocular reflex, 5
- Voronoi cells, 466
- Volterra models, 762
- Weak learning model, 358
- Weierstrass theorem, 249
- Weight-sharing, 28, 89
- Weighted norm, 280
- Wiener filters, 127–128
- Willshaw-von der Malsburg's model, 146
- Winner-takes-all neuron, 58
- Woodbury's equality, see Matrix inversion lemma
- XOR problem, 175–178, 252, 260–261, 282–284, 335–337
- Z-transform, 637

