

References

- D. Ackley, G. E. Hinton and J. Sejnowski, A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147–169, 1985.
- A. Agarwal and H. Daumé III, A geometric view of conjugate priors. *Machine Learning*, 81, 99–113, 2010.
- M. Aizerman, E. Braverman and L. Rozonoer, Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25, 821–837, 1964.
- M. Akahira and K. Takeuchi, *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*, Springer LN in Statistics, vol. 7, 1981.
- S. Akaho and K. Takabatake, Information geometry of contrastive divergence. In *Information Theory and Statistical Learning*, 3–9, 2008.
- M. S. Ali and S. D. Silvey, A general class of coefficients of divergence of one distribution from another. *Journal of Royal Statistical Society, B*, 28, 131–142, 1966.
- S. Amari, On some primary structures of non-Riemannian plasticity theory. *RAAG Memoirs*, 3, D-IX, 99–108, 1962.
- S. Amari, A geometrical theory of moving dislocations. *RAAG Memoirs*, 4, D-XVII, 153–161, 1968.
- S. Amari, Theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, 16, 299–307, 1967.
- S. Amari, Neural theory of association and concept-formation. *Biological Cybernetics*, 26, 175–185, 1977.
- S. Amari, Differential geometry of curved exponential families—curvature and information loss. *Annals of Statistics*, 10, 357–385, 1982.
- S. Amari, Finsler geometry of non-regular statistical models. *RIMS Kokyuroku* (in Japanese), Non-Regular Statistical Estimation, Ed. M. Akahira, 538, 81–95, 1984.
- S. Amari, *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics, 28, Springer, 1985.
- S. Amari, Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections and divergence. *Mathematical Systems Theory*, 20, 53–82, 1987.
- S. Amari, Information geometry of the EM and em algorithms for neural networks. *Neural Networks*, 8, 1379–1408, 1995.
- S. Amari, Natural gradient works efficiently in learning. *Neural Computation*, 10, 251–276, 1998.
- S. Amari, Superefficiency in blind source separation. *IEEE Transactions on Signal Processing*, 47, 936–944, 1999.
- S. Amari, Estimating functions of independent component analysis for temporally correlated signals. *Neural Computation*, 12, 2083–2107, 2000.

- S. Amari, Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, 47, 1701–1711, 2001.
- S. Amari, Integration of stochastic models by minimizing α -divergence. *Neural Computation*, 19, 2780–2796, 2007.
- S. Amari, α -divergence is unique, belonging to both f -divergence and Bregman divergence classes. *IEEE Transactions on Information Theory*, 55, 11, 4925–4931, 2009.
- S. Amari, Information geometry of positive measures and positive-definite matrices: Decomposable dually flat structure. *Entropy*, 16, 2131–2145, 2014.
- S. Amari and J. Armstrong, Curvature of Hessian manifolds, *Differential Geometry and its Applications* 33, 1–12, 2014.
- S. Amari and J-F. Cardoso, Blind source separation—Semiparametric statistical approach. *IEEE Transactions on Signal Processing*, 45, 2692–2700, 1997.
- S. Amari, A. Cichocki and H. Yang, A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems* (Eds. M. Mozer et al.), 8, 757–763, 1996.
- S. Amari and M. Kawanabe, Information geometry of estimating functions in semi-parametric statistical models. *Bernoulli*, 3, 29–54, 1997.
- S. Amari, S. Ikeda and H. Shimokawa, Information geometry of α -projection in mean field approximation. In M. Opper and D. Saad (Eds), *Advanced Mean Field Methods: Theory and Practice*, 241–257. MIT Press, 2001.
- S. Amari, K. Kurata and H. Nagaoka, Information geometry of Boltzmann machines. *IEEE Transactions on Neural Networks*, 3, 260–271, 1992.
- S. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Society and Oxford University Press, 2000.
- S. Amari, H. Nakahara, S. Wu and Y. Sakai, Synchronous firing and higher-order interactions in neuron pool. *Neural Computation*, 15, 127–142, 2003.
- S. Amari and A. Ohara, Geometry of q -exponential family of probability distributions. *Entropy*, 13, 1170–1185, 2011.
- S. Amari, A. Ohara and H. Matsuzoe, Geometry of deformed exponential families: Invariant, dually flat and conformal geometry. *Physica A*, 391, 4308–4319, 2012.
- S. Amari, H. Park and K. Fukumizu, Adaptive method of realizing natural gradient learning for multilayer perceptrons. *Neural Computation*, 12, 1399–1409, 2000.
- S. Amari, H. Park and T. Ozeki, Singularities affect dynamics of learning in neuromanifolds. *Neural Computation*, 18, 1007–1065, 2006.
- S. Amari and S. Wu, Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12, 783–789, 1999.
- S. Amari and M. Yukawa, Minkovskian gradient for sparse optimization. *IEEE Journal of Selected Topics in Signal Processing*, 7, 576–585, 2013.
- D. Arthur and S. Vassilvitskii, k -means++: The advantages of careful seeding, *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035, 2007.
- K. Arwini and C. T. J. Dodson, *Information Geometry*. Springer, 2008.
- N. Ay, An information-geometric approach to a theory of pragmatic structuring. *Annals of Probability*, 30, 416–436, 2002.
- N. Ay, Information geometry on complexity and stochastic interaction. *Entropy*, 17, 2432–2458, 2015.
- N. Ay, J. Jost, H. V. Lê and L. Schwachhöfer, *Information Geometry and Sufficient Statistics*. [arXiv:1207.6736](https://arxiv.org/abs/1207.6736), 2013.
- N. Ay and S. Amari, A novel approach to canonical divergences within information geometry. *Entropy*, 17, 8111–8129, 2015.
- N. Ay and A. Knauf, Maximizing multi-information. *Kybernetika*, 42, 517–538, 2006.
- N. Ay, E. Olbrich, N. Bertschinger and J. Jost, A geometric approach to complexity. *Chaos*, 21, 037103, 2011.
- D. Balduzzi and G. Tononi, Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Computational Biology*, 4, e1000091, 2008.

- A. Banerjee, S. Merugu, I. Dhillon and J. Ghosh, Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6, 1705–1749, 2005.
- N. Barkai, H. S. Seung, and H. Sompolinsky. On-line learning of dichotomies. *Advances in Neural Information Processing Systems*, 7, 303–310, 1995.
- O. E. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*. Wiley, 1978.
- A. B. Barrett and A. K. Seth, Practical measures of integrated information for time-series data. *PLoS Computational Biology*, 7, e1001052, 2011.
- M. Basseville, Divergence measures for statistical data processing—An annotated bibliography. *Signal Processing*, 93, 621–633, 2013.
- A. Beck and M. Teboulle, Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31, 167–175, 2003.
- J. M. Begun, W. J. Hall, W. M. Huang and J. A. Wellner, Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics*, 11, 432–452, 1983.
- A. J. Bell and T. Sejnowski, An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129–1159, 1995.
- P. J. Bickel, C. A. J. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1994.
- J.-D. Boissonnat, F. Nielsen and R. Nock, Bregman Voronoi diagrams. *Discrete and Computational Geometry*, 44, 281–307, 2010.
- L. Bregman, The relaxation method of finding a common point of convex sets and its applications to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7, 200–217, 1967.
- R. Brockett, Some geometric questions in the theory of linear systems. *IEEE Transactions on Automatic Control*, 21, 449–455, 1976.
- R. Brockett, Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. *Linear Algebra and its Applications*, 146, 79–91, 1991.
- A. Bruckstein, D. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51, 34–81, 2009.
- J. Burbea and C. R. Rao, On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28, 489–495, 1982.
- W. Byrne, Alternating minimization and Boltzmann machine learning. *IEEE Transactions on Neural Networks*, 3, 612–620, 1992.
- O. Calin and C. Udriste, *Geometric Modeling in Probability and Statistics*. Springer, 2013.
- L. L. Campbell, An extended Chentsov characterization of a Riemannian metric. *Proceedings of American Mathematical Society*, 98, 135–141, 1986.
- E. J. Candes, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* 59, 1207–1223, 2006.
- E. J. Candes and M. B. Walkin, An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25, 21–30, 2008.
- J.-F. Cardoso and B. H. Laheld, Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44, 3017–3030, 1996.
- J.-F. Cardoso and A. Souloumiac, Jacobi angles for simultaneous diagonalization. *SIAM Journal on Mathematical Analysis and Applications*, 17, 161–164, 1996.
- A. Cena and G. Pistone, Exponential statistical manifold. *Annals of Institute of Statistical Mathematics*, 59, 27–56, 2007.
- T. Chen and S. Amari, Unified stabilization approach to principal and minor components extraction algorithms. *Neural Networks*, 14, 1377–1387, 2001.
- T. P. Chen, S. Amari and Q. Lin, A unified algorithm for principal and minor components extraction. *Neural Networks*, 11, 3, 385–390, 1998.
- S. S. Chen, D. L. Donoho and M. A. Saunders, Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computation*, 20, 33–61, 1998.
- N. N. Chentsov, *Statistical Decision Rules and Optimal Inference*, AMS, 1982 (originally published in Russian, Nauka, 1972).

- H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Annals of Mathematical Statistics*, 23, 493–507, 1952.
- H. Choi, S. Choi, A. Kataké and Y. Choe, Parameter learning for α -integration. *Neural Computation*, 25, 1585–1604, 2013.
- J. Choi and A. P. Mullhaupt, Kahlerian information geometry for signal processing. *Entropy*, 17, 1581–1605, 2015.
- A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. John Wiley, 2002.
- A. Cichocki and S. Amari, Families of α -, β - and γ -divergences: flexible and robust measures of similarities. *Entropy*, 12, 1532–1568, 2010.
- A. Cichocki, S. Cruces and S. Amari, Generalized α - β divergences and their application to robust nonnegative matrix factorization. *Entropy*, 13, 134–170, 2011.
- A. Cichocki, S. Cruces and S. Amari, Log-determinant divergences revisited: α - β and γ log-det divergences. *Entropy*, 17, 2988–3034, 2015.
- A. Cichocki, R. Zdunek, A. H. Phan and S. Amari, *Nonnegative Matrix and Tensor Factorizations*. John Wiley and Sons, UK, 2009.
- C. Cortes and V. Vapnik, Support-vector networks. *Machine Learning*, 20, 273–297, 1995.
- F. Cousseau, T. Ozeki and S. Amari, Dynamics of learning in multilayer perceptrons near singularities. *IEEE Transactions on Neural Networks*, 19, 1313–1328, 2008.
- F. Critchley, P. K. Marriott and M. Salmon, Preferred point geometry and statistical manifolds. *Annals of Statistics*, 21, 1197–1224, 1993.
- I. Csizsár, Information-type measures of difference of probability distributions and indirect observation. *Studia Scientiarum Mathematicarum Hungarica*, 2, 229–318, 1967.
- I. Csizsár, Information measures: A critical survey. in *Proceedings of 7th Conference on Information Theory*, Prague, Czech Republic, 83–86, 1974.
- I. Csizsár, Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Annals of Statistics*, 19, 2032–2066, 1991.
- I. Csizsár and G. Tusnady, Information geometry and alternating minimization procedure. In E. F. Dedewicz, et. al. (Eds.), *Statistics and Decision*, 205–237, Oldenburg Verlag, 1984.
- Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli and Y. Bengio, Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. [arXiv:1406.2572](https://arxiv.org/abs/1406.2572), NIPS, 2014.
- A. P. Dawid, The geometry of proper scoring rules. *Annals of Institute of Statistical Mathematics*, 59, 77–93, 2007.
- A. P. Dawid, S. Lauritzen and M. Parry, Proper local scoring rules on discrete sample spaces. *Annals of Statistics*, 40, 593–608, 2012.
- A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. *Journal Royal Statistical Society*, B, 39, 1–38, 1977.
- S. Dhillon and J. A. Tropp, Matrix nearness problems with Bregman divergences. *SIAM Journal on Matrix Analysis and Applications*, 29, 1120–1146, 2007.
- D. L. Donoho, Compressed sensing. *IEEE Transactions on Information Theory*, 52, 1289–1306, 2006.
- D. L. Donoho and Y. Tsaig, Fast solution of L_1 -norm minimization problems when the solution may be sparse. *IEEE Transaction on Information Theory*, 54, 4789–4812, 2008.
- A. Edelman, A. A. Arias and S. T. Smith, The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20, 303–353, 1998.
- B. Efron, Defining the curvature of a statistical problem (with application to second order efficiency). *Annals of Statistics*, 3, 1189–1242, 1975.
- B. Efron, T. Hastie, I. Johnstone and R. Tibshirani, Least angle regression. *Annals of Statistics*, 32, 407–499, 2004.
- S. Eguchi, Second order efficiency of minimum contrast estimators in a curved exponential family. *Annals of Statistics*, 11, 793–803, 1983.
- S. Eguchi, O. Komori and A. Ohara, Duality of maximum entropy and minimum divergence. *Entropy*, 16, 3552–3572, 2014.

- M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.
- Y. Eldar and G. Kutyniok, *Compressed Sensing*. Cambridge University Press, 2012.
- Y. Freund and R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. *Journal Computer and Systems Sciences*, 55, 119–139, 1997.
- H. Fujisawa and S. Eguchi, Robust parameter estimation with a small bias against heavy contamination. *Journal Multivariate Analysis*, 99, 2053–2081, 2008.
- A. Fujiwara and S. Shuto, Hereditary structure in Hamiltonians: Information geometry of Ising spin chains. *Physics Letters A*, 374, 911–914, 2010.
- K. Fukumizu, Likelihood ratio of unidentifiable models and multilayer neural networks. *Annals of Statistics*, 31, 833–851, 2003.
- K. Fukumizu, Exponential manifold by reproducing kernel Hilbert spaces. In *Algebraic and Geometric Methods in Statistics* (P. Gibilisco, E. Riccomagno, M.-P. Rogantin and H. Winn Eds.), 291–306, Cambridge University Press, 2009.
- K. Fukumizu and S. Kuriki, Statistics of Singular Models. *Frontiers in Statistical Sciences*, 7, Iwanami, 2004 (in Japanese).
- S. Furuichi, An axiomatic characterization of a two-parameter extended relative entropy. *Journal of Mathematical Physics*, 51, 2010.
- P. Gibilisco and G. Pistone, Connections on non-parametric statistical manifolds by Orlicz space geometry: infinite-dimensional analysis. *Quantum Probabilities and Related Topics*, 1, 325–347, 1998.
- M. Girolami and B. Calderhead, Riemannian manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of Royal Statistical Society, B-73*, 123–214, 2011.
- V. P. Godambe, *Estimating Functions*. Oxford University Press, 1991.
- M. Grasselli, Dual connections in nonparametric classical information geometry. *Annals of Institute of Statistical Mathematics*, 62, 873–896, 2010.
- I. Grondman, L. Buşoniu, G.A.D. Lopes and R. Babuška, A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, 42, 1291–1307, 2012.
- P. D. Grünwald and A. P. Dawid, Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, 32, 1367–1433, 2004.
- N. Hansen and A. Ostermeier, Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9, 159–195, 2001.
- G. H. Hardy, J. E. Littlewood and G. Polya, *Inequalities* (2nd ed.). Cambridge: Cambridge University Press, 1952.
- K. V. Harsha and K. S. S. Moosath, F -geometry and Amari's α -geometry on a statistical manifold. *Entropy*, 16, 2472–2487, 2014.
- M. Hayashi and S. Watanabe, Information geometry approach to parameter estimation in Markov chains. *IEEE Transactions on Information Theory*, 2014.
- M. Henmi and R. Kobayashi, Hooke's law in statistical manifolds and divergence. *Journal Nagoya Mathematical*, 159, 1–24, 2000.
- G. E. Hinton, Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14, 1771–1800, 2002.
- G. E. Hinton and E. R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science*, 313, 504–507, 2006.
- Y. Hirose and F. Komaki, An extension of least angle regression based on the information geometry of dually flat spaces. *Journal of Computational and Graphical Statistics*, 19, 1007–1023, 2010.
- S. W. Ho and R. W. Yeung, On the discontinuity of the Shannon information measures. *IEEE Transactions on Information Theory*, 55, 5362–5374, 2009.
- A. Honkela, T. Raiko, M. Kuusela, M. Törnio and J. Karhunen, Approximate Riemannian conjugate gradient learning for fixed-form variational Bayes. *Journal of Machine Learning Research*, 11, 3235–3268, 2010.

- A. Hyvärinen, Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6, 695–709, 2005.
- A. Hyvärinen, Some extensions of score matching. *Computational Statistics & Data Analysis*, 51:2499–2512, 2007.
- A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*. John Wiley, 2001.
- T. Ichimori, On rounding off quotas to the nearest integers in the problem of apportionment methods. *SIAM Letters*, 3, 21–24, 2011.
- S. Ikeda, T. Tanaka and S. Amari, Stochastic reasoning, free energy, and information geometry. *Neural Computation*, 16, 1779–1810, 2004a.
- S. Ikeda, T. Tanaka and S. Amari, Information geometry of turbo and low-density parity-check codes. *IEEE Transactions on Information Theory*, 50, 1097–1114, 2004b.
- M. Ishikawa, Structural learning with forgetting. *Neural Networks*, 9, 509–521, 1996.
- R. A. Jacobs, M. I. Jordan, S. J. Nolwan and G. E. Hinton, Adaptive mixtures of local experts. *Neural Computation*, 3, 79–87, 1991.
- A. T. James, The variance information manifold and the function on it. *Multivariate Statistical Analysis*, Ed. P. K. Krishnaiah, Academic Press, 157–169, 1973.
- H. Jeffreys, *Theory of Probability*, 1st ed. Clarendon Press, 1939.
- H. Jeffreys, An invariant form for the prior probability in estimation problems. *Proceedings of Royal Society of London, Series A, Mathematical and Physical Sciences*, 186, 453–461, 1946.
- H. Jeffreys, *Theory of Probability*, 2nd ed. Oxford University Press, 1948.
- K. Jeong, M. Yukawa and S. Amari, Can critical-point paths under l_p -regularization ($0 < p < 1$) reach the sparsest least square solutions?. *IEEE Transactions on Information Theory*, 60, 2960–2968, 2014.
- J. Jiao, T. M. Courtade, A. No, K. Venkat and T. Weissman, Information measure: The curious case of the binary alphabet. *IEEE Transactions on Information Theory*, 60, 7616–7626, 2015.
- S. Kakade, A natural policy gradient. In *Advances in Neural Information Processing*, 14, 1531–1538, 2001.
- S. Kakihara, A. Ohara and T. Tsuchiya, Information geometry and interior-point algorithms in semi-definite programs and symmetric cone programs. *Journal of Optimization Theory and Applications*, DOI [10.1007/s10957-012-0189-9](https://doi.org/10.1007/s10957-012-0189-9), 2012.
- T. Kanamori, T. Takenouchi, S. Eguchi and N. Murata, Robust loss function for boosting. *Neural Computation*, 19, 2183–2244, 2007.
- K. Kanatani, Statistical optimization and geometric inference in computer vision. *Philosophical Transactions of Royal Society of London, Ser. A*, 356, 1303–1320, 1998.
- Y. Kano, Beyond third-order efficiency. *Sankhya*, 59, 179–197, 1997.
- Y. Kano, More higher order efficiency. *Journal of Multivariate Analysis*, 67, 349–366, 1998.
- R. Karakida, M. Okada and S. Amari, Analyzing feature extraction by contrastive divergence learning in RBM. *NIPS Workshop on Deep Learning*, 2014.
- R. Karakida, M. Okada and S. Amari, Dynamical analysis of contrastive divergence learning. *Restricted Boltzmann machines with Gaussian visible units*, To appear, 2016.
- R. E. Kass and P. Vos, *Geometrical Foundations of Asymptotic Inference*. Wiley, 1997.
- A. Kim, J. Park S. Park and S. Kang, Impedance learning for robotic contact tasks using natural actor-critic algorithm. *IEEE Transactions on Systems, Man and Machine*, B39, 433–443, 2010.
- J. Kivinen and M. K. Warmuth, Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132, 1–63, 1997.
- G. Kniadakis and A. Scarfone, A new one parameter deformation of the exponential function. *Physica A*, 305, 69–75, 2002.
- K. Kumon and S. Amari, Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference. *Proceedings of Royal Society of London, A* 387, 429–458, 1983.
- M. Kumon, A. Takemura and K. Takeuchi, Conformal geometry of statistical manifold with application to sequential estimation. *Sequential Analysis*, 30, 308–337, 2011.
- T. Kurose, Dual connections and affine geometry. *Mathematische Zeitschrift*, 203, 115–121, 1990.

- T. Kurose, On the divergence of 1-conformally flat statistical manifolds. *Tohoku Mathematical Journal*, 46, 427–433, 1994.
- T. Kurose, Conformal-projective geometry of statistical manifolds. *Interdisciplinary Information Sciences*, 8, 89–100, 2002.
- S. Lauritzen, *Graphical Models*. Oxford University Press, 1996.
- H. V. Lê, Statistical manifolds are statistical models. *Journal of Geometry*, 84, 83–93, 2005.
- G. Lebanon and J. Lafferty, Boosting and maximum likelihood for exponential models. In *Advances in Neural Information Processing Systems (NIPS)*, 14, 2001.
- D. D. Lee and S. Seung, Algorithms for nonnegative matrix factorization. *Nature*, 401, 788–791, 1999.
- C. Lin and J. Jiang, Supervised optimizing kernel locality preserving projection with its application to face recognition and palm biometrics. Submitted, 2015.
- M. Liu, B. C. Vemuri, S. Amari and F. Nielsen, Shape retrieval using hierarchical total Bregman soft clustering. *IEEE Transactions on Pattern Analysis and Machine Learning*, 34, 2407–2419, 2012.
- L. Malagò, M. Matteucci and G. Pistone, Natural gradient, fitness modelling and model selection: A unifying perspective. *IEEE Congress on Evolutionary Computation*, 486–493, 2013.
- L. Malagò and G. Pistone, Combinatorial optimization with information geometry: Newton method. *Entropy* 16, 4260–4289, 2014.
- P. Marriotti, On the local geometry of mixture models. *Biometrika*, 89, 77–93, 2002.
- P. Marriotti and M. Salmon, *Applications of Differential Geometry to Econometrics*. Academic Press, 2011.
- J. Martens, New perspectives on the natural gradient method. [arXiv:1412.1193](https://arxiv.org/abs/1412.1193), 2015.
- J. Martens and R. Grosse, Optimizing neural networks with Kronecker-factored approximate curvature. [arXiv:1503.05671](https://arxiv.org/abs/1503.05671), 2015.
- R. J. Martin, A metric for ARMA processes. *IEEE Transactions on Signal Processing*, 48, 1164–1170, 2000.
- T. Matumoto, Any statistical manifold has a contrast function—On the C3-functions taking the minimum at the diagonal of the product manifold. *Hiroshima Mathematical Journal*, 23, 327–332, 1993.
- Y. Matsuyama, The α -EM algorithm: Surrogate likelihood maximization using α -logarithmic information measures. *IEEE Transactions on Information Theory*, 49, 692–706, 2003.
- H. Matsuzoe, On realization of conformally-projectively flat statistical manifolds. *Hokkaido Mathematical Journal*, 27, 409–421, 1998.
- H. Matsuzoe, Geometry of contrast functions and conformal geometry. *Hokkaido Mathematical Journal*, 29, 175–191, 1999.
- H. Matsuzoe, J. Takeuchi and S. Amari, Equiaffine structures on statistical manifolds and Bayesian statistics. *Differential Geometry and Its Applications*, 24, 567–578, 2006.
- J. Milnor, On the concept of attractor. *Communications of Mathematical Physics*, 99, 177–195, 1985.
- M. Minami and S. Eguchi, Robust blind source separation by β -divergence. *Neural Computation*, 14, 1859–1886, 2004.
- K. Miura, M. Okada and S. Amari, Estimating spiking irregularities under changing environments. *Neural Computation*, 18, 2359–2386, 2006.
- T. Morimoto, Markov processes and the H -theorem. *Journal of Physical Society of Japan*, 12, 328–331, 1963.
- T. Morimura, E. Uchibe, J. Yoshimoto and K. Doya, A generalized natural actor-critic algorithm. In *Advances in Neural Information Processing Systems*, 22, MIT Press, 1312–1320, 2009.
- R. Morioka and K. Tsuda, Information geometry of input-output table. Technical Report IEICE, 110, 161–168, 2011 (in Japanese).
- N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, Information geometry of U -boost and Bregman divergence. *Neural Computation*, 16, 1432–1481, 2004.
- M. K. Murray and J. W. Rice, *Differential Geometry and Statistics*. Chapman Hall, 1993.

- H. Nagaoka and S. Amari, Differential geometry of smooth families of probability distributions. Technical Report METR 82-7, University of Tokyo, 1982.
- H. Nakahara and S. Amari, Information-geometric measure for neural spikes. *Neural Computation*, 14, 2269–2316, 2002.
- H. Nakahara, S. Amari and B. Richmond, A comparison of descriptive models of a single spike train by information-geometric measure. *Neural Computation*, 18, 545–568, 2006.
- J. Naudts, *Generalized Thermostatistics*. Springer, 2011.
- A. Nemirovski and D. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley, 1983.
- Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms*. SIAM Publications, 1993.
- Y. Nesterov and M. Todd, On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, 2, 333–361, 2002.
- N. J. Newton, An infinite-dimensional statistical manifold modeled on Hilbert space. *Journal of Functional Analysis*, 263, 1661–1681, 2012.
- J. Neyman and E. L. Scott, Consistent estimates based on partially consistent observation. *Econometrica*, 16, 1–32, 1948.
- F. Nielsen and R. Nock, On the χ -square and higher-order χ -distances for approximating f -divergences. *IEEE Signal Processing Letters*, 21, 10–13, 2014.
- R. Nock and F. Nielsen, Bregman divergences and surrogates for learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 2048–2059, 2009.
- R. Nock, F. Nielsen and S. Amari, On conformal divergences and their population minimizers. *IEEE Transactions on Information Theory*, accepted, 2015.
- K. Nomizu and T. Sasaki, *Affine Differential Geometry*. Oxford University Press, 1994.
- A. Ohara, Information geometric analysis of an interior point method for semidefinite programming. In O. Barndorff-Nielsen and E. Jensen Eds, *Geometry in Present Day Science*, World Scientific, 49–74, 1999.
- A. Ohara, Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Physics Letters, A*, 370, 184–193, 2007.
- A. Ohara and S. Amari, Differential geometric structures of stable state feedback systems with dual connections. *Kybernetika*, 30, 369–386, 1994.
- A. Ohara and S. Eguchi, Group invariance of information geometry on q -Gaussian distributions induced by beta-divergence. *Entropy*, 15, 4732–4747, 2013.
- A. Ohara, H. Matsuzoe and S. Amari, Conformal geometry of escort probability and its applications. *Modern Physics Letters B*, 26, 10, 1250063, 2012.
- M. Oizumi, L. Albantakis and G. Tononi, From phenomenology to the mechanism of consciousness: Integrated information theory 3.0. *PLoS Computational Biology*, 10, e1003588, 2014.
- M. Oizumi, S. Amari, T. Yanagawa, N. Fujii and N. Tsuchiya, Measuring integrated information from the decoding perspective. [arXiv:1505.04368](https://arxiv.org/abs/1505.04368) [q-bio.NC], To appear in *PLoS Computational Biology*, 2015.
- M. Oizumi, M. Okada and S. Amari, Information loss associated with imperfect observation and mismatched decoding. *Frontiers in Computational Neuroscience*, 5, 1–13, 2011.
- M. Oizumi, N. Tsuchiya and S. Amari, A unified framework for information integration based on information geometry. Submitted, 2016.
- E. Oja, A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15, 267–273, 1982.
- E. Oja, Principal components, minor components, and linear neural networks. *Neural Networks*, 5, 927–935, 1992.
- I. Okamoto, S. Amari and K. Takeuchi, Asymptotic theory of sequential estimation: Differential-geometrical approach. *Annals of Statistics*, 19, 961–981, 1991.
- T. Okatani and K. Deguchi, Easy calibration of a multi-projector display system. *International Journal of Computer Vision*, 2009.

- Y. Ollivier, Riemannian metric for neural networks I: Feedforward networks. *Information and Inference*, 4, 108–153, 2015, DOI [10.1093/imaiai/iav006](https://doi.org/10.1093/imaiai/iav006).
- H. Park, S. Amari and K. Fukumizu, Adaptive natural gradient learning algorithms for various stochastic models. *Neural Networks*, 13, 755–764, 2000.
- M. Parry, A. P. Dawid and S. Lauritzen, Proper local scoring rule. *Annals of Statistics*, 40, 561–592, 2012.
- J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.
- J. Peters and S. Schaal, Natural actor-critic. *Neurocomputing*, 71, 1180–1190, 2008.
- G. Pistone, Examples of the application of nonparametric information geometry to statistical physics. *Entropy*, 15, 4042–4065, 2013.
- G. Pistone and M. P. Rogantin, The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli*, 5, 721–760, 1999.
- G. Pistone and C. Sempì, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Annals of Statistics*, 23, 1543–1561, 1995.
- C. R. Rao, Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37, 81–91, 1945.
- C. R. Rao, Efficient estimates and optimum inference procedures in large samples. *Journal of Royal Statistical Society, B*, 24, 46–72, 1962.
- G. Raskutti and S. Mukherjee, The information geometry of mirror descent. *IEEE Transactions on Information Theory*, 61, 1451–1457, 2015.
- J. Rauh, Finding the maximizers of the information divergence from an exponential family. *IEEE Transactions on Information Theory*, 57, 3236–3247, 2011.
- N. Ravishanker, E. L. Melnik and C. Tsai, Differential geometry of ARMA models. *Journal of Time Series Analysis*, 11, 259–274, 1990.
- A. Rényi, On measures of entropy and information, in *Proc. 4th Symposium on Mathematical Statistics and Probability Theory*, Berkeley, CA, 1, 547–561, 1961.
- F. Rosenblatt, *Principles of Neurodynamics*. Spartan, 1962.
- N. L. Roux, P.-A. Manzagol and Y. Bengio, Topmoumoute online natural gradient algorithm. In *Advances in Neural Information Processing Systems*, 17, 849–856, 2007.
- D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning representations by back-propagating errors. *Nature*, 323, 533–536, 1986.
- R. E. Schapire, Y. Freund, P. Bartlett and W. S. Lee, Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, 26, 1651–1686, 1998.
- J. Schmidhuber, Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85–117, 2015.
- B. Scholkopf, *Support Vector Learning*. Oldenbourg, 1997.
- J. A. Schouten, *Ricci Calculus*. Springer, 1954.
- J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- H. Shima, *The Geometry of Hessian Structures*. World Scientific, 2007.
- S. Shinomoto, K. Shima and J. Tanji, Differences in spiking patterns among cortical neurons. *Neural Computation*, 15, 2823–2842, 2003.
- P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing*, 1, 194–281, MIT Press, 1986.
- A. Soriano and L. Vergara, Fusion of scores in a detection context based on alpha integration. *Neural Computation*, 27, 1983–2010, 2015.
- S. M. Stigler, The epic story of maximum likelihood. *Statistical Science*, 22, 598–620, 2007.
- T. Takenouchi and S. Eguchi, Robustifying AdaBoost by adding the naive error rate. *Neural Computation* 16, 767–787, 2004.
- T. Takenouchi, S. Eguchi, N. Murata and T. Kanamori, Robust boosting algorithm against mislabeling in multiclass problems. *Neural Computation*, 20, 1596–1630, 2008.

- J. Takeuchi, Geometry of Markov chains, finite state machines and tree models. Technical Report of IEICE, 2014.
- J. Takeuchi, T. Kawabata and A. Barron, Properties of Jeffreys mixture for Markov sources. *IEEE Transactions on Information Theory*, 41, 643–652, 2013.
- K. Tanabe, Geometric method in nonlinear programming. *Journal of Optimization Theory and Applications*, 30, 181–210, 1980.
- T. Tanaka, Information geometry of mean field approximation. *Neural Computation*, 12, 1951–1968, 2000.
- M. Taniguchi, Higher-Order Asymptotic Theory for Time Series Analysis. Springer Lecture Notes in Statistics, 68, 1991.
- P. S. Thomas, W. Dabney, S. Mahadeven and S. Giguere, Projected natural actor-critic. In *Advances in Neural Information Processing Systems*, 26, 2013.
- R. Tibshirani, Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society, Series B*, 58, 267–288, 1996.
- G. Tononi, Consciousness as integrated information: a provisional manifest. *Biological Bulletin*, 215, 216–242, 2008.
- F. Topsoe, Information-theoretical optimization techniques. *Kybernetika*, 15, 8–27, 1979.
- C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52, 479–487, 1988.
- C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, 2009.
- K. Uohashi, α -conformal equivalence of statistical submanifolds. *Journal of Geometry*, 75, 179–184, 2002.
- V. N. Vapnik, *Statistical Learning Theory*. John Wiley, 1998.
- B. C. Vemuri, M. Liu, S. Amari and F. Nielsen, Total Bregman divergence and its applications to DTI analysis. *IEEE Transactions on Medical Imaging*, 30, 475–483, 2011.
- R. F. Vigelis, and C. C. Cavalcante, On ϕ -families of probability distributions. *Journal of Theoretical Probabilities*, 26, 870–884, 2013.
- P. Vos, A geometric approach to detecting influential cases. *Annals of Statistics*, 19, 1570–1581, 1991.
- J. Wada, A divisor apportionment method based on the Kolm-Atkinson social welfare function and generalized entropy. *Mathematical Social Sciences*, 63, 243–247, 2012.
- M. J. Wainwright and M. I. Jordan, Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1, 1–305, 2008.
- S. Watanabe, Algebraic geometrical methods for hierarchical learning machines. *Neural Networks*, 14, 1409–1060, 2001.
- S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, 2009.
- S. Watanabe, Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular statistical learning theory. *Journal of Machine Learning Research*, 11, 3571–3591, 2010.
- H. Wei and S. Amari, Dynamics of learning near singularities in radial basis function networks. *Neural Networks*, 21, 989–1005, 2008.
- H. Wei, J. Zhang, F. Cousseau, T. Ozeki and S. Amari, Dynamics of learning near singularities in layered networks. *Neural Computation*, 20, 813–843, 2008.
- P. Williams, S. Wu and J. Feng, Two scaling methods to improve performance of the support vector machine. In *Support Vector Machines: Theory and Applications*, Ed. L. Wang, 205–218, Springer, 2005.
- D. Wu, Parameter estimation for α -GMM based on maximum likelihood criterion. *Neural Computation*, 21, 1776–1795, 2009.
- S. Wu and S. Amari, Conformal transformation of kernel functions: A data-dependent way to improve support vector machine classifiers. *Neural Processing Letters*, 15, 59–67, 2002.

- S. Wu, S. Amari and H. Nakahara, Population coding and decoding in a neural field: A computational study. *Neural Computation*, 14, 999–1026, 2002.
- L. Xu, Least mean square error reconstruction principle for self-organizing neural nets. *Neural Networks*, 6, 627–648, 1993.
- Z. Xu, X. Chang, F. Xu and H. Zhang, $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on Neural Networks and Learning Systems*, 23, 1013–1027, 2012.
- S. Yi, D. Wierstra, T. Schaul and J. Schmidhuber, Stochastic search using the natural gradient. *ICML Proceedings of the 26th Annual International Conference on Machine Learning*, 1161–1168, 2009.
- J. S. Yedidia, W. T. Freeman and Y. Weiss, Generalized belief propagation. In T. K. Leen, T. G. Dietrich and V. Tresp (Eds.), *Advances in Neural Information Processing Systems*, 13, 689–695, MIT Press, 2001.
- A. Yuille, CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation. *Neural Computation*, 14, 1691–1722, 2002.
- A. L. Yuille and A. Rangarajan, The concave-convex procedure. *Neural Computation*, 15, 915–936, 2003.
- M. Yukawa and S. Amari, l_p -regularized least squares ($0 < p < 1$) and critical path. *IEEE Transactions on Information Theory*, 62, 1–15, 2016.
- J. Zhang, Divergence function, duality and convex analysis. *Neural Computation*, 16, 159–195, 2004.
- J. Zhang, From divergence function to information geometry: Metric, equiaffine and symplectic structures. *Geometry Symposium, Japan Mathematical Society, Proceedings*, 47–62, 2011.
- J. Zhang, Nonparametric information geometry: From divergence function to referential-representational biduality on statistical manifolds. *Entropy*, 15, 5384–5418, 2013.
- J. Zhang, On monotone embedding in information geometry. *Entropy*, 17, 4485–4499, 2015.
- J. Zhao, H. Wei, C. Zhang, W. Li, W. Guo and K. Zhang, Natural gradient learning algorithms for RBF networks. *Neural Computation*, 27, 481–505, 2015.
- H. Y. Zhu and R. Rohwer, Bayesian invariant measurements of generalization. *Neural Processing Letters*, 2, 28–31, 1995.

Index

Symbols

- $(\alpha\text{-}\beta)$ -divergence, 100
- (α, β) -divergence, 94
- $(\alpha\text{-}\beta)$ -log-det divergence, 101
- α -divergence, 58, 67, 72
- α -expert machine, 84
- α -family of probability distributions, 81
- α -function, 57
- α -geodesic, 75
- α -geometry, 136
- α -integration, 82
- α -mean, 77
- α -projection theorem, 76
- α -Pythagorean theorem, 76
- β -divergence, 95
- χ -divergence, 91
- χ -escort distribution, 91
- χ -exponential family, 90
- e -affine parameter, 38
- e -condition, 257
- e -flat, 38
- e -geodesic, 38
- e -parallel transport, 202
- em algorithm, 28
- (F, G, H) -structure, 104
- f -divergence, 54
- γ -divergence, 102
- k -cut, 144
- k -means, 234
- k -sparse, 336
- κ -exponential family, 89
- L_0 -norm, 338
- L_1 -norm, 338
- m -affine parameter, 38
- m -condition, 257
- m -flat, 38
- m -geodesic, 38
- m -parallel transport, 202
- m -projection, 46, 252
- ϕ -center of cluster, 233
- Φ -function method, 245
- q -divergence, 85
- q -entropy, 85
- q -escort geometry, 92
- q -exponential, 85
- q -exponential family, 86, 89
- q -free energy, 87
- q -logarithm, 85
- q -metric, 88
- (ρ, τ) -structure, 104
- U -divergence, 95
- (u, v) -divergence, 92
- (u, v) -structure, 99

A

- Absolute-value-based Hessian natural gradient, 286
- Active set, 340
- Adaptive learning method, 294
- Adaptive natural gradient learning, 293
- Affine connection, 112
- Affine coordinate system, 18
- Affine flat structure, 19
- Akaike information criterion, 312
- Alternating minimization algorithm, 27
- Amari–Chentsov structure, 134
- Amari–Chentsov tensor, 134
- Ancillary submanifold, 169
- Ancillary tangent subspace, 201
- ARMA model, 218
- AR model, 217
- Asymptotic theory of hypothesis testing, 175
- Auto-correlation coefficients, 221

Auto-regression model, 217

B

Back-propagation learning, 281

Barrier function, 345

Basis vectors, 20

Bayesian duality, 266

Bayesian posterior distribution, 266

Belief propagation, 249

Blow-down technique, 309

Boltzmann machine, 181, 268

Boosting, 261

Bregman divergence, 13

C

Canonical divergence, 138

Canonical parameter, 32

Central limit theorem, 60

Chernoff divergence, 242

Chernoff information, 242

Clique, 250

Clustering, 231

Clustering algorithm, 234

Coarse graining, 53

Cocktail party problem, 323

Coefficient of proportionality, 191

Coefficients of affine connection, 113

Conformal transformation, 91

Conformal transformation of a kernel, 249

Conjugate priors, 267

Consistent estimator, 168

Contrastive divergence, 273

Convex-concave computational procedure,
249

Convex function, 12

Convex programming, 345

Coordinate system, 4

Coordinate transformation, 4

Covariant derivative, 117

Cramér–Rao bound, 166

Cramér–Rao theorem, 166

Critical region, 300

Critical slowdown, 305

Cubic tensor, 115

Cumulant generating function, 32

D

Decomposable divergence, 55

Deep learning, 292, 296

Deformed exponential family, 89

Divergence, 9

Dual affine structure, 19

Dual connections, 131

Dual convex function, 17

Dual geodesic, 19

Dually flat manifold, 137

E

Efficient, 173

Efficient score, 194

Einstein summation convention, 20

Eliminating singularity, 299

EM algorithm, 179

Embedding curvature, 129, 349

Ergodic time series, 215

Error covariance matrix, 166

Escort probability distribution, 88

Estimating function, 197

Estimator, 165

Euler–Schouten curvature, 129

Exponential family, 31

F

First-order asymptotic theory, 173

Fisher information matrix, 33

Foliation, 145

Free energy, 32

G

Game, 349

Game-divergence, 350

Game-score, 349

Gaussian kernel, 247

Gaussian mixture model, 180

Gaussian RBM, 275

Generalization error, 280

Generalized inverse, 337

Generalized Pythagorean theorem, 24

Geodesic, 19, 117

Graph Laplacian, 356

Graphical model, 250

H

Hidden variable, 179

Higher-order asymptotic theory of estimation, 173

Higher-order correlations, 149

Higher-order cumulants, 326

Hyvärinen divergence, 354

Hyvärinen score, 353

I

Independent component analysis, 322
 Information integration, 150
 Information monotonicity, 52
 Inner product, 23
 Input–output analysis, 157
 Instantaneous loss, 280
 Integrated information, 152
 Integration of weak machines, 261
 Invariance criterion, 51
 Invariant divergences, 52
 Invariant Riemannian metrics, 52

K

Kernel exponential family, 42
 Kernel function, 246
 Killing metric, 329
 KL-divergence, 71, 220
 Kronecker-factored approximate curvature, 293
 Kullback–Leibler (KL) divergence, 11

L

Large deviation, 60
 Large deviation theorem, 61
 Learning constant, 294
 Least angle regressions, 343
 Least equiangle theorem, 342
 Legendre transformation, 16
 Levi–Civita connection, 113, 125
 Linear machine, 242
 Linear system, 215
 Loss of information by data reduction, 185

M

Machine learning, 231
 MA model, 218
 Manifold, 3
 Margin, 243
 Maximum entropy, 223
 Maximum entropy principle, 45
 Maximum likelihood estimator, 48
 Mean field approximation, 254
 Metric affine connection, 125
 Milnor attractor, 310
 Minimum description length, 312
 Minimum entropy, 224
 Minkovskian gradient, 343
 Minor subspace, 317
 Mirror descent method, 289
 Misspecified model, 186

Mixed coordinate system, 144
 Mixture family, 37
 Moving-average model, 218
 Multilayer perceptron, 292, 296

N

Natural gradient, 283
 Natural gradient learning method, 284
 Natural parameter, 32
 Natural policy gradient, 288
 Negative entropy, 33
 Neyman–Scott problem, 191
 Non-holonomic coordinate system, 329
 Non-negative matrix factorization, 333
 Nuisance parameter, 191
 Nuisance tangent subspace, 201

O

Observed point, 47
 Observed submanifold, 180
 On-line learning, 281
 Overlapping singularity, 299

P

Parallel transport, 22, 118
 Parameter of interest, 191
 Plateau, 302
 Plateau phenomena, 308
 Policy natural gradient, 284
 Polynomial kernel, 247
 Positive-definite symmetric matrix, 96
 Power spectrum, 217
 Principal component, 317
 Principal component analysis, 315
 Principal subspace, 317
 Prior distribution, 266
 Projection theorem, 25, 143

R

RAS transformation, 160
 RC curvature, 119
 Reinforcement learning, 287
 Restricted Boltzmann machine, 268
 Riemann–Christoffel curvature tensor, 119
 Riemannian connection, 113
 Riemannian geometry, 109
 Riemannian gradient, 283
 Riemannian metric, 19
 Riemannian structure, 10
 Robust cluster center, 238

S

Saddle-free Newton method, 286
Scale problem, 191
Score function, 111
Semi-definite programming, 346
Shape parameter, 212
Singular point, 301
Singular prior, 313
Singular statistical models, 311
Singular structure, 296
Soft clustering, 236
Solution path, 341
Sparse vector, 336
Standard estimating function, 332
Standard f -divergence, 56
Stiefel manifold, 320
Stochastic descent learning method, 281
Stochastic relaxation, 286
Submanifold, 126
Sufficient statistic, 52
Super efficiency, 332
Support vector, 244

Support vector machine, 242

System complexity, 152

T

Tangent space, 19, 109
Tangent subspace of interest, 201
Temporal firing pattern, 211
Tensor, 114
Time series, 215
Total Bregman divergence, 238
Total least squares, 196
Training error, 280
Transfer function, 217

U

Unidentifiability, 298

V

Voronoi diagram, 234