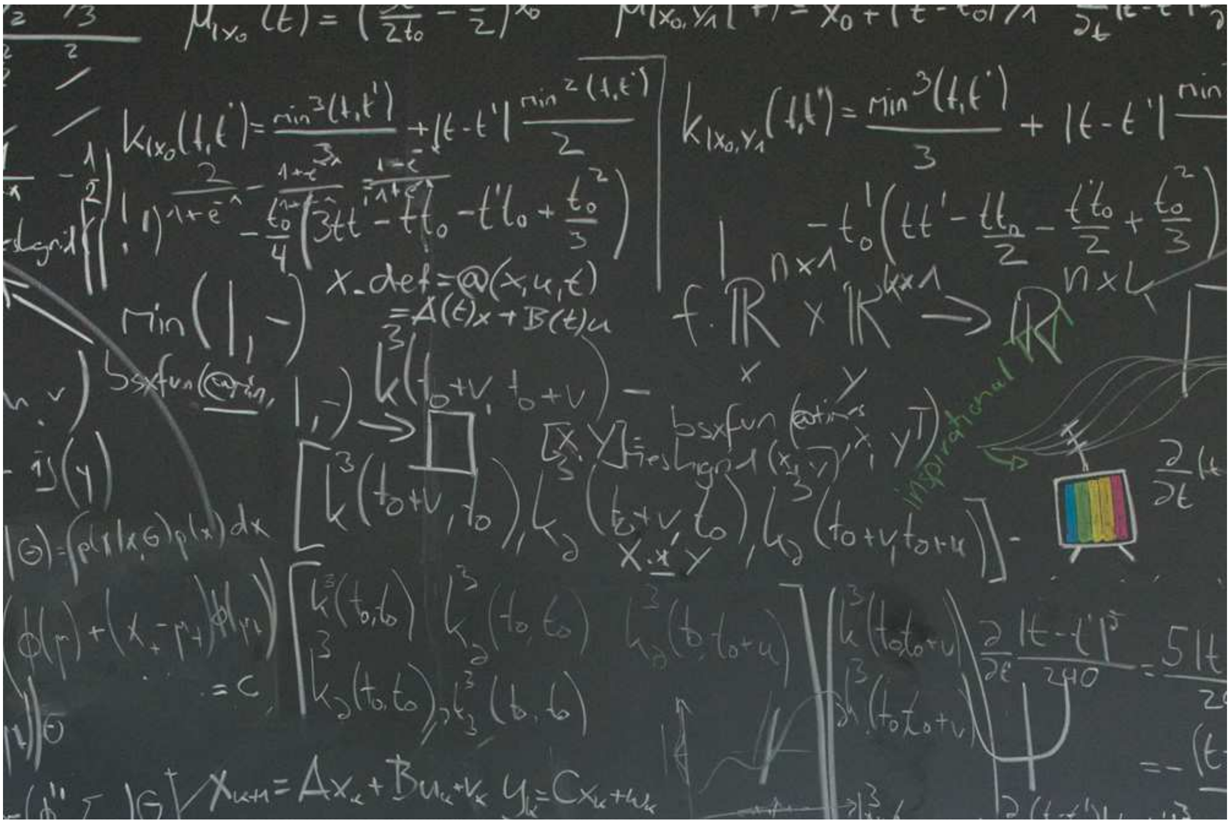


Empirical Inference

Report for the
Scientific Advisory Board
April 18-20, 2016

Max Planck Institute for Intelligent Systems, Stuttgart • Tübingen

MAX-PLANCK-GESELLSCHAFT

Contents

1 EMPIRICAL INFERENCE

1.1 Research Overview

The problems studied in the department[1] can be subsumed under the heading of *empirical inference*, i.e., inference performed on the basis of empirical data. This includes statistical learning, but also the inference of causal structures from statistical data, leading to models that provide insight into the underlying mechanisms, and make predictions about the effect of interventions. Likewise, the type of empirical data can vary, ranging from biological measurements (e.g., in neuroscience) to astronomical observations. We are conducting theoretical, algorithmic, and experimental studies to try and understand the problem of empirical inference.

The department was started around statistical learning theory and kernel methods. It has since broadened its set of inference tools to include a stronger component of Bayesian methods, including graphical models with a recent focus on issues of causality. In terms of the inference tasks being studied, we have moved towards tasks that go beyond the relatively well-studied problem of supervised learning, such as semi-supervised learning or transfer learning. Finally, we have continuously striven to analyze challenging datasets from biology, astronomy, and other domains, leading to the inclusion of several application areas in our portfolio. No matter whether the applications are done in the department or in collaboration with external partners, considering a whole range of applications helps us *study principles and methods of inference*, rather than inference applied to one specific problem domain.

The most competitive publication venues in empirical inference are NIPS (Neural Informa-

[1] The department of EI previously belonged to the MPI for biological Cybernetics. The present extended evaluation covers the period since the last departmental evaluation as part of that institute, i.e., since 2009.

tion Processing Systems), ICML (International Conference on Machine Learning), UAI (Uncertainty in Artificial Intelligence), and for theoretical work, COLT (Conference on Learning Theory). The presence at these conferences makes us one of the top international machine learning labs. In addition, we sometimes submit our work to the leading application oriented conferences in neighboring fields including computer vision (ICCV, ECCV, CVPR) and data mining (KDD, ICDM, SDM), as well as to specialized journals. Our work has earned us a number of awards, including best paper prizes at COLT 2003, NIPS 2004, COLT 2005, COLT 2006, ICML 2006, ALT 2007, DAGM 2008, CVPR 2008, ISMB 2008, ECCV 2008, NIPS 2008, NIPS 2009, UAI 2010, ICINCO 2010, DAGM 2011, IROS 2012, NIPS 2013, CIP 2014, and honorable mentions at IROS 2008, NIPS 2009, DAGM 2009, KDD 2010, ECML/PKDD 2011, and IROS 2012.

Below, we briefly describe our research directions in more detail. We begin with a high-level overview of the broad directions, followed by short descriptions of a selection of projects.

Theoretical studies, algorithms, and applications often go hand in hand. For instance, it may be the case that someone working on a specific application will develop a customized algorithm that turns out to be of independent theoretical interest. Such serendipity is a desired side effect caused by interaction across groups and research areas, for instance during our frequent departmental talks. It concerns cross-fertilization of methodology (e.g., kernel independence measures used in causal inference), the transfer of algorithmic developments or theoretical insights to application domains (e.g., causal inference in neuroscience), or the combination of different application areas (e.g., using methods of computational photography for MR motion correction in neuroscience).

The linear organization of the text does not permit an adequate representation of all these connections. Below, we have opted for an organization of the material that devotes individual sections to our main application areas (computational imaging, robot learning, neuroscience, and bioinformatics), and that comprises four methodological sections, on learning algorithms, causal inference, probabilistic inference, and statistical learning theory. We begin with the latter.

Statistical Learning Theory A machine learning algorithm is given training data and tries to learn a model that is well-suited to describe the data and that can be used to make predictions. The goal of statistical learning theory is to assess to which extent such algorithms can be successful in principle. The general approach is to assume that the training data have been generated by an unknown random source, and to develop mathematical tools to analyze the performance of a learning algorithm in statistical terms: for example, by bounding prediction errors ("generalization bounds") or by analyzing large sample behavior and convergence of algorithms on random input ("consistency"). Over the last decade, statistical learning theory has made significant progress in the analysis of supervised learning problems such as classification and regression. The department has made various contributions to this area, including two recent NIPS orals (2011, 2012), where we prove that regressors can escape the curse of dimensionality if the intrinsic dimensionality of the data is low [546], and provide methods to increase performance by introducing gradient weights [481]. We have, however, gradually shifted our attention to areas of machine learning where statistical learning theory is less well developed. These include settings that incorporate both labeled and unlabeled data, such as semi-supervised or transductive learning, active learning and transfer learning, as well as unsupervised learning tasks such as clustering, manifold learning, graph-based learning, and the emerging field of complex network science. Our goal is to contribute statistical foundations to these exciting new challenges where pioneering work can be done.

Active learning exploits structure and information in unlabeled data to reduce label supervision by requesting labels only for a small set of points from a large pool. We provided novel statistical guarantees for cluster-based active learning procedures [340] and a first formal analysis of the use of active learning for learning under covariate shift [341].

We studied theoretical aspects of transductive learning [342], introduced a new complexity measure of function classes called Permutational Rademacher Complexity, which is better suited for transductive problems than other popular analogues, and used it to derive data-dependent generalization bounds for transductive algorithms.

One focus in the area of random geometric graphs was on the limit behavior of distance functions as the size of the graph increases. We showed that several of the well-known distance functions on graphs, such as the shortest path distance or the commute distance, converge to undesired limit functions that cannot be justified for applications. As an alternative, we constructed a family of p-resistance distances that shows an interesting phase transition behavior and, in certain regimes of p, does not suffer from those problems.

We conclude with two projects connected to other departmental research themes. For reinforcement learning, we derived PAC-Bayesian concentration inequalities for multiple simultaneously evolving and interdependent martingales [138]. We applied the inequalities to analyze importance-weighted sampling and exploration-exploitation trade-off in multi-armed bandits with side information [536], yielding adaptation to the relevant side information and run time improvements. For causal inference, we developed an approach posing cause-effect inference as binary classification of probability measures. We proposed a computationally efficient algorithm based on kernel mean embeddings achieving state of the art performance, and provided a risk bound for the approach [356].

Learning Algorithms Learning algorithms based on kernel methods have enjoyed considerable success in a wide range of supervised learning tasks, such as regression and classification. One reason for the popularity of these approaches is that they solve difficult non-parametric problems by mapping data points into high dimensional spaces of features, specifically reproducing kernel Hilbert spaces (RKHSes), in which linear algorithms can be brought to bear, leading to solutions taking the form of kernel expansions [487].

In recent years, we have developed kernel mappings of *probability distributions* to RKHSes [142]. For sufficiently rich feature spaces, these maps can be shown to be injective [253], i.e., they represent probability distributions without any loss of information. This property is profoundly linked to questions of optical imaging, by viewing a Fraunhofer imaging process as a particular case of kernel mapping [423].

Using such mappings, we can compare samples from two distributions, to determine whether they are distinguishable to a statistically significant degree [142]. This comprises the problem of independence testing as a special case, where we determine the difference between the embedding of a joint distribution and that of the product of the marginals.

They can be generalized to yield embeddings of conditional distributions, which are a core building block of a number of learning algorithms. Conditional distribution embeddings can be used to perform Bayesian inference on graphical models [556, 558, 602]. A second application is in reinforcement learning, where conditional expectations are used to update the state of an agent given its observations and actions [462]. A final application of conditional mean embeddings is the challenging task of testing for conditional dependence [390, 519]. This is of interest for finding which covariates contain redundant information, but also when performing causal inference, by providing a generalization of the well-known PC algorithm by Glymour and Spirtes that can incorporate nonlinear interactions.

Although kernel methods were originally associated with statistical learning theory, our work has helped move them towards probabilistic approaches. This includes our generalization of SVMs to the case where the inputs are not point observations, but nontrivial probability measures [415, 478]. This allows us to model uncertainty of the inputs. It also provides a generalization of SVMs towards multi-scale kernel expansions that naturally arise as a consequence of nontrivial input measures, as well as a generalization of SVMs towards multi-layer systems.

Kernel means are used throughout the field nowadays, and it is thus of significant interest to study their estimation properties. This led us to generalizations of classical results on Stein shrinkage estimation [35, 395].

Recently, we have started exploiting kernel means in the context of probabilistic programming, proposing an elegant way of representing the distribution of arbitrary arithmetic functions of random variables. Being a kernel method, this approach is applicable also to non-vectorial data types and accordingly holds promise for generalizing standard computing operations to distributions over data types [43]. Further details on our work on learning algorithms, including for the new field of network science, is included in the project reports below.

Causal inference The detection and use of statistical dependences form the core of statistics and machine learning. In recent years, machine learning methods have enabled us to perform rather accurate prediction, often based on large training sets, for complex nonlinear problems that not long ago would have appeared completely random. However, in many situations we would actually prefer a causal model to a purely predictive one; i.e., a model that might tell us that a specific variable (say, whether or not a person smokes) is not just statistically associated with a disease, but causal for the disease.

Pearl's graphical approach to causal modeling generalizes Reichenbach's common cause principle and characterizes the observable statistical (conditional) independences that a causal structure should entail. Many causal inference methods build on these independences to infer causal graphs from data. This "graphical models" approach to causal inference has several weaknesses that we try to address in our work: it only can infer causal graphs up to Markov equivalence, it does not address the hardness of conditional independence testing, and it usually does not worry about the complexity of the underlying functional regularities that generate statistical dependences in the first place. Our work in this field is characterized by the following three aspects:

1. We usually work in terms of structural equation models (SEMs) or functional causal models (FCMs), i.e., we do not take statistical dependences as primary, but rather study mechanistic models which give rise to such dependences. In FCMs, each variable is modeled as a deterministic function of its direct causes and some noise variable N, e.g., $Y = f(X, Z, N)$; all noise variables are assumed to be jointly independent. FCMs do not only allow us to model observational distributions; one can also use them in order to model what happens under interventions (e.g., gene knockouts or randomized studies). Our work in this domain includes the use of restricted function classes such as additive noise models, which solve the previously open problems of cause-effect inference [84, 624] and allow the detection of confounders in certain cases [625]. It also includes functional models where the functions are implemented by Turing machines and the implied dependences are not statistical ones, but characterized in terms of Kolmogorov complexity. We have developed a theory of graphical models for such models, analogous to the known probabilistic approach. It uses algorithmic mutual information rather than the usual statistical notion of mutual information, and it can be derived from a computing model (using Turing machines) just like graphical models can be derived from FCMs. Since it is based on algorithmic complexity of bit strings, it is in principle also applicable to individual (non-statistical) observations [234, 590].

2. Viewed from an FCM perspective, the crucial assumption of the graphical approach to causality is statistical independence of all noise terms. Intuitively, it is clear that as the noises propagate through the graph, they pick up dependences due to the graph structure, hence the assumption of initial independence allows us to tease out properties of that structure. We believe, however, that much can be gained by considering a more *general independence assumption* related to notions of invariance and autonomy of causal mechanisms. Here, the idea is that causal mechanism are autonomous entities of the world that (in the generic case) do not depend on each other, and changing (or intervening on) one of them often leaves the remaining ones invariant.

In a first study, we were able to show that if the effect is a deterministic invertible function of the cause which satisfies a suitable independence condition relative to the cause distribution, then this independence condition is provably violated between the effect and the inverse function [134, 562]. We have used the general independence assumption to connect the problems of covariate shift and transfer learning to causality, arguing that the causal conditionals (or mechanisms) should often be invariant under domain changes [360, 470].[2] We have also exploited it to develop methods for inferring causal graphs based on this kind of invariance [20].

3. This leads to the third characteristic aspect of our work on causality. Wherever possible, we attempt to establish connections to machine learning, and indeed we believe that some of

[2] M. Rojas-Carulla, B. Schölkopf, R. Turner, J. Peters. Causal transfer in machine learning. preprint http://arxiv.org/abs/1507.05333. 2016.

the hardest problems of machine learning (such as those of domain adaptation and transfer) are best addressed using causal thinking. We have shed light on the conditions under which semi-supervised learning (SSL) works, showing that knowledge of the underlying causal structure allows us to predict whether SSL can possibly work [24, 470]. Likewise, we believe that machine learning approaches are useful in advancing the field of causality. Some examples include the kernel-based tests for (conditional) independence developed in our lab [390, 519], our work on consistency of cause-effect inference [400], and the application of kernel mean embeddings to learn and analyze cause-effect-inference as a classification problem [356].

Our lab has played a major role in putting causal inference on the agenda of the machine learning community, and we expect that causal inference will have practical implications for many scientific inference problems, especially where interventions are not feasible (e.g., in astronomy [349] and neuroscience [14, 34]). It touches statistics, econometrics, and philosophy, and it constitutes one of the most exciting field for conceptual basic research in machine learning today.

Probabilistic Inference The probabilistic formulation of inference—conditioning probability measures encoding prior assumptions on observed data by multiplying with a likelihood describing the data's assumed generative process—remains one of the main research streams within machine learning. One of our main themes in this field has been nonparametric inference on function spaces using Gaussian process models [240, 371, 543]. We also produced widely used software for Gaussian process regression [249] and inference in generalized linear models [130].

The central computational bottleneck in Bayesian models is the marginalization of latent variables. This can be computationaly demanding, so approximate inference routines reducing computational complexity are a major research theme. This includes work on variational models, expectation propagation [498], and custom-built approximations for specific applications [232, 452, 651], as well as general integration routines [396] constructed *de novo*.

Members of the department have also been interested in uses of probabilistic inference for reinforcement learning and control [16, 404].

A broad new research theme has emerged around *probabilistic numerics*, the description of computation itself as probabilistic inference. Although this work was started in the department and several group members are still department members, currently we are not discussing it in further detail. More on this direction can be found in the project section on probabilistic numerics, and the separate section on the Emmy Noether Research Group that has arisen from it.

Computational Imaging Our work in this domain focuses on how to recover an image from corrupted measurements by modeling the distortion process and solving the resulting ill-posed inverse problem using image priors to restore the most likely image explaining our observations. We studied two major types of distortions: blur and noise. The latter included learning approaches using neural networks, some of them beating the state of the art BM3D algorithm [450, 539], as well as a collaboration with a department of our Stuttgart site [52].

For the purpose of deblurring, we developed an online method for *multi-frame blind deconvolution* [620] and generalized it to be able to super-resolve and to be robust against saturated pixels. To extend the applicability of deconvolution algorithms to spatially varying blur, we proposed the *Efficient Filter Flow* (EFF) model [570], and combined it successfully with our online blind deconvolution method [620]. This paved the way to consider *camera shake*, a major source of non-uniform blur in photography. While it had previously been modeled as spatially invariant, real-world shake leads to spatially variant blur that we modeled using EFF, improving upon the state-of-the-art [597]. By incorporating physical constraints into the shake model, we significantly decreased running time and improved reconstruction quality [537, 718].[3]

Non-uniform blur also results from *lens aberrations*, usually minimized through sophisticated lens design. Nonetheless, even professional lenses exhibit residual aberrations, resulting in smoothly varying blur that tends to get worse in the image corners. We developed meth-

[3] These methods have led to a start-up involving several department members.

ods that automatically reduce lens aberrations from images, with or without a priori knowledge of the lens blur [472, 535]. More recently, we cast this into a data-driven regression framework, leading to what is probably the current state of the art [320]. This has the potential to change future lens designs by shifting the focus from expensive hardware solutions towards a combination of hardware and sophisticated software.

Two relevant application areas of computational imaging are astronomy and medical imaging, and we briefly mention some of our work in that field, starting with the latter. Motivated by our work on camera shake removal, we assayed whether similar ideas could be applied to *magnetic resonance imaging (MRI) motion correction*. The setup is different from regular camera shake: MR images are taken by recording trajectories in Fourier space ("k-space"), and a moving object will blur the different frequencies in a way that depends on the order in which the k-space is scanned. By carefully modeling the distortion process and imposing a sharpness prior on the resulting image, we were able to recover excellent images from sequences that would be rendered useless for standard MR reconstruction algorithms. This work applies not only for translational movements but also for rotational ones [98]. A second line of research, pursued in collaboration with the University of Tübingen's Laboratory for Preclinical Imaging, applies machine learning and image processing methods in combined PET-MR systems. We have developed a solution to the problem of PET attenuation correction, which was licensed to Siemens and won a best paper award of the field's flagship journal [103, 205].

In astronomy, our work initially focused on high-resolution imaging using deconvolution [184], but we applied our expertise also to other astronomical data analysis problems. Recent highlights include an approach to combine heterogeneous imaging data by voting methods [361], as well as causal modeling methods for light curves that have led to the discovery of a number of previously unknown exoplanets [38, 349].

Robot Learning Research in robotics and artificial intelligence has lead to the development of complex robots ranging from anthropomorphic arms to complete humanoids. In order to be meaningfully applied in human-inhabited environments, robots need to possess a variety of physical abilities and skills. However, programming such skills is a labor- and time-intensive task which requires a large amount of expert knowledge. In particular, it often involves transforming intuitive concepts of motions and actions into formal mathematical descriptions and algorithms. To overcome such difficulties, we use imitation learning to teach robots new motor skills [238]. A human demonstrator first provides one or several examples of the skill. Information recorded through motion capture or physical interaction is used by the robot to automatically generate a controller that can replicate the movements. This step can be accomplished using modern machine learning techniques. Imitation learning allows robots to learn the observed behavior and have a good starting point for self-improvement. This so called self-improvement of the task can help the robot adapt the learned movement to the characteristics of its own body or the requirements of the current context. Hence, even if the examples presented by humans are not optimal, the robot can still use them to bootstrap its behavior. In our lab, imitation learning has already been used to teach complex motor skills to various kinds of robots. This includes skills such as grasping of novel objects [239, 574, 575], robot table tennis [48, 581], ball-in-the-cup [178], throwing & catching [144, 467, 499], and tether ball [455, 465]. Our work on grasping won the ICINCO 2010 Best Paper Award. New machine learning methods that can acquire motor skills faster are developed. The goal of this research is to have intelligent robots that can autonomously enlarge their repertoire of skills by observing or interacting with human teachers.

However, supervised learning is not always sufficient for motor learning problems, partly because an expert teacher or idealized version of the behavior is often not available. Accordingly, one of our goals is the development of reinforcement learning methods which can handle the dimensionality of humanoid robots and generate actions for seven or more degrees of freedom. Efficient reinforcement learning for continuous states and actions is essential for robotics and control [683]. For high-dimensional state and action spaces, it is often easier to directly learn policies without estimating accurate system models. The resulting algorithms are parametric policy search algorithms inspired by expectation-maximization methods and can be employed for motor primitive learning [144, 178, 465]. Our resulting approach [465] has won the IROS 2012 Best Cognitive Paper, and was both IROS 2012 Best Student Paper Finalist and Best Paper Award Finalist.

Our work on reinforcement learning (RL) also comprises algorithmic and theoretical developments that are of independent interest. In this context, significant progress has been made regarding new information theoretic RL approaches [218, 508, 586], hierarchical RL approaches [455], as well as structured approaches [463]. We also studied theoretical foundations for Bayesian RL approaches [540] and the analysis of reinforcement learning algorithms in the bandit setting [536] and within inverse reinforcement learning [508].

Machine Learning in Neuroscience The neurosciences present some of the steepest challenges to machine learning. Nearly always there is a very high-dimensional input structure — particularly relative to the number of exemplars, since each data point is usually gathered at a high cost. To avoid overfitting, inference must thus make considerable use of domain knowledge. Relevant regularities are often subtle, the rest being made up of noise that may be of much larger magnitude (often composed largely of the manifestations of other neurophysiological processes, besides the ones of interest). In finding generalizable solutions, one usually has to contend with a high degree of variability, both between individuals and across time, leading to problems of covariate shift and non-stationarity.

One specific neuroscientific application area in which we have a long-standing interest is that of brain-computer interfacing, or BCI (see page 30). This research aims to construct systems that replace and rehabilitate lost communication and motor skills in patient populations. The contribution of machine learning to BCI is in developing, refining, and applying algorithms to improve the accuracy with which neural signals are decoded. Our algorithmic developments ([13, 186, 187, 283, 560]) as well as improvements in experimental methodology (for example, as described in [137, 223, 298]) are significant contributions towards making clinical BCI systems a reality.

In addition to the advancement of methods for brain decoding, further progress in this field is contingent on a better understanding of the neural basis of BCI control. Research has largely focused on neural signals that provide information on a subject's intent, neglecting those processes that determine whether a subject is in a state of mind suitable for operating a BCI. In collaboration with the Institute of Medical Psychology and Behavioral Neurobiology at the University of Tübingen, we have studied the neuro-physiological causes of good and poor BCI-control in healthy subjects and patient populations. This work has provided insights into the relation of attention networks and BCI-performance [149, 174], as recognized by the annual *International BCI Research Award 2011*, and has led to a new class of cognitive BCI paradigms for patients in late stages of amyotrophic lateral sclerosis (ALS) [81, 346].

Our interest in interfacing with the brain is not confined to communication. We are further exploring the use of BCI systems as assistive devices in stroke rehabilitation. Online decoding of movement intent in severely disabled patients can be employed to deliver congruent haptic feedback through robotic devices, thereby re-establishing the disrupted sensorimotor feedback loop. This synchronization may support cortical reorganization, hopefully resulting in enhanced recovery [183]. We have developed a BCI-controlled robotic system in collaboration with the Neurosurgery Department at the University of Tübingen, and demonstrated the feasibility of real-time haptic feedback on movement intent in healthy subjects and stroke patients [194].

We have also designed machine learning techniques to assist with the interpretation of experimental brain data (see page 31). Unsupervised learning tools were designed to automatically identify key events in large neural recordings, and were an integral part of several high-impact publications on memory-triggering events [25, 160]. Finding causal relationships between neural processes is also of particular interest to neuroscientists, but hard to address in living neural systems due to ethical and practical concerns. We have been developing causal inference tools adapted to these challenges [46, 174, 224, 353], and this work led to significant neuroscientific findings such as the role of traveling electrical waves in routing sensory information across the visual cortex [34]. Machine learning also provided theoretical insights into biological learning by showing that observed properties of synaptic plasticity are optimal for neurons to learn under metabolic constraints imposed by limited energetic resources [95, 473].

We conclude this section with a short summary of our contributions to cognitive science and vision research, performed in collaboration with Felix Wichmann, professor of computational neuroscience at the local University Tübingen and part-time member of our department. In one line of work, we use machine learning to uncover the behaviorally relevant features observers use in complex perceptual tasks: given high-dimensional sensory input, which are the features that sensory systems base their computations on? We have applied our methods successfully to visual saliency [290], gender discrimination [242], and auditory tone-in-noise detection [148]. We have also published a discussion of the application of kernel methods in the flagship journal of cognitive science [289]. A second focus is the development of a predictive image-based model of spatial vision. We integrated the large psychophysical literature on simple detection and discrimination experiments and proposed a model based on maximum-likelihood decoding of a population of model neurons [108].

Bioinformatics In the past, the department was active in several branches of computational biology. While these activities have been significantly reduced, they overlapped with the reporting period and will thus briefly be included.

Genetics has witnessed an impressive increase in datasets that contain detailed information both about the genetic properties of individuals and their appearance properties, that is about their *genotype* and *phenotype*. Large genetics consortia around the world are recording single-nucleotide polymorphisms (SNPs), i.e., single genome bases that differ between individuals, for large sets of individuals. The standard technique is then to compute which of these 10^5 to 10^7 genome positions correlate most strongly with the disease status. Due to the enormous computational costs that an exhaustive scan for, say, 10^{14} candidate pairs would entail, these studies usually ignore any interactive effect between genomic loci, although these are believed to be important in many diseases. We have developed new engineering approaches and mathematical results to tackle this fundamental computational problem in Statistical Genetics [513].

Another recent area in Computational Biology is Molecular Genomics, in which new sequencing technologies enable an unprecedented wealth of data for exploring the structure of the genome and of individual genes. We have developed several tools to detect genes within the genome, and to establish a fine-grained picture of gene structure and of the impact of sequence variation on gene expression levels [312, 313, 650].

1.2 Selected Research Projects

Learning theory

Machine learning algorithms are designed to yield accurate predictions on future data by generalizing from past observations. The main goal of learning theory is to analyze their statistical and computational properties, and to provide performance guarantees on the new data. To do so, it poses these tasks in a rigorous mathematical framework and deals with them under various assumptions on the data-generating process.

PAC-Bayesian analysis (Probably Approximately Correct) provides guarantees on the generalization ability of randomized predictors within the classical PAC framework. We derived general PAC-Bayesian inequalities which allow to control the concentration of weighted averages of possibly uncountably many simultaneously evolving and interdependent martingales [138]. We applied these inequalities to **contextual bandits** providing a new regret bound and learning algorithm [536].

We have studied the relations and differences between **Vapnik-Chervonenkis** (VC) and **Popper's dimensions**. While both VC and Popper's dimensions aim to quantify the process of falsification, the first focuses on passive falsification based on examples provided by nature, whereas the latter focuses on active falsification through experiment design. We have shown that concepts that are related to Popper's notion of falsifiability occur in the domain of query-learning and derived relations between Popper's, exclusion, and VC-dimension [677].

We have analyzed how the performance of learning algorithms **adapts to local structure** in the data, which often helps when the ambient dimension is huge. We established that certain regression procedures, such as kernel and k-NN regression, adapt to a locally low intrinsic dimension [439, 546]. Further, we have shown how to improve regression performance by adapting the learning process to differing variability in various coordinates of the function to be estimated [481]. For this, we proposed a novel, consistent estimator for gradient norms. We also developed new data-structures for regression in a streaming setting [440].

We showed the first statistical consistency result (under surprisingly general distributional conditions) for additive noise methods in **causal inference** [400]. Further, we posed the problem of cause-effect inference as binary classification of probability measures, proposed a computationally efficient algorithm based on kernel mean embeddings achieving state-of-the-art performance, and provided a risk bound for the approach [356] .

In **transductive learning** the learner observes labeled training and unlabeled test points with the final goal of correctly labeling the test points. We have introduced a new complexity measure of function classes called Permutational Rademacher Complexity, argued that it is better suited for analysis of transductive learning than other popular measures of complexity, and derived a general risk bound for transductive algorithms based on it [342].

Active learning exploits structure and information in unlabeled data to save label supervision. The goal is to ask for only a small subset of the data to be labeled while maintaining good performance. We provided novel statistical guarantees for a large class of practical cluster-based active learning procedures [340]. We further showed that an algorithmic technique from margin-based active learning yields a computationally efficient learner for linear separators under conditions that had earlier been shown to allow for fast statistical learning rates (Massart noise condition) [339].

We also provided a first formal analysis of the use of active learning for **domain adaptation**, that is, to use a learner's active querying ability to adapt to changes in the data generation. We proposed a new active learning strategy and proved that it consistently learns in situations where no passive learning algorithm is consistent, while automatically adapting its amount of label queries to what is needed due to the underlying data shift [341].

More information: https://ei.is.tuebingen.mpg.de/project/learning-theory

Random geometric graphs

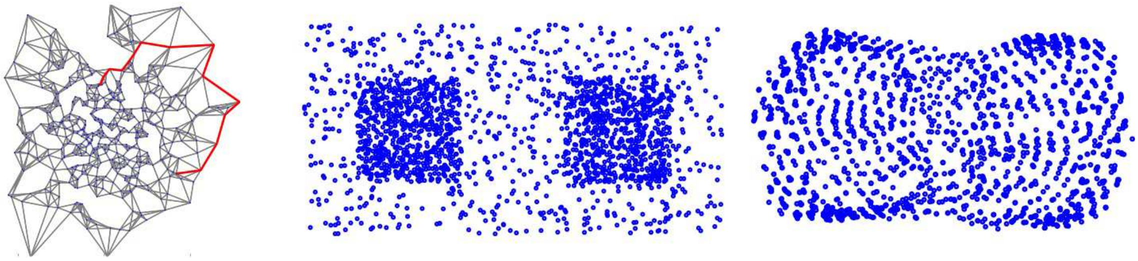


Figure 1.1: Left: Shortest path (in red) in an unweighted k-nearest neighbor graph on a random set of points. The path is far from the straight line between the two end points. Middle and right: a toy data set (middle) and its Isomap reconstruction based on the shortest path distance in the unweighted kNN graph. The reconstrution does not preserve the density information of the original data set.

Random geometric graphs are built by first sampling a set of points from some underlying distribution, and then connecting each point to its k nearest neighbors. In this project we investigated the behavior of distance functions on random geometric graphs when the sample size n goes to infinity (and the connectivity parameter k scales appropriately).

It is well known that in graphs where the edges are suitably weighted according to their Euclidean lengths, the **shortest path distance** converges to the underlying Euclidean distance. However, it turned out this is not the case for unweighted kNN graphs [461]. In this case, the shortest path distance converges to a distance function that is weighted by the underlying density and takes wide detours to avoid high density regions (see left figure above). In machine learning applications, this behavior of the shortest path distance can be highly misleading. As an example, consider the Isomap algorithm and the data set shown in the middle figure. If we build an unweighted kNN graph based on this data and apply Isomap to recover the point configuration, we get the figure on the right. Obviously, it is grossly distorted and cannot serve as a faithful representation of the original data.

The **commute distance** (aka **resistance distance**) between vertex u and v is defined as the expected time it takes the natural random walk starting in vertex u to travel to vertex v and back. It is widely used in machine learning because it supposedly satisfies the following, highly desirable property: Vertices in the same cluster of the graph have a small commute distance, whereas vertices in different clusters of the graph have a large commute distance to each other. We studied the behavior of the commute distance as the number of vertices in the graph tends to infinity [599], proving that the commute distance between two points converges to a trivial quantity that only takes into account the degree of the two vertices. Hence, all information about cluster structure gets lost when the graph is large enough.

To alleviate this shortcoming, we proposed the family of p**-resistances** [497]. For $p = 1$ it reduces to the shortest path distance, for $p = 2$ it coincides with the resistance distance, and for $p \to \infty$ it is related to the minimal s-t-cut in the graph. The family shows an interesting phase transition: there exist two critical thresholds p^* and p^{**} such that if $p < p^*$, then the p-resistance depends on meaningful global properties of the graph, whereas if $p > p^{**}$, it only depends on trivial local quantities and does not convey any useful information. In particular, the p-resistance for $p = p^*$ nicely reveals the cluster structure.

More information: https://ei.is.tuebingen.mpg.de/project/distance-functions-on-random-geometric-graphs

Kernel methods

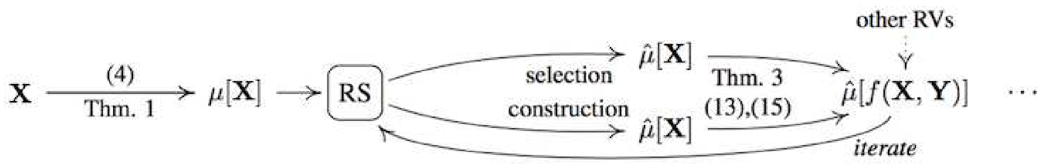


Figure 1.2: Overview of the application of kernel mean embedding in [43]. Together with the reduced set (RS) techniques to limit the complexity of the RKHS expansion, the kernel mean embedding is used to approximate the embedding of the functional of random variables $Z = f(X, Y)$.

A Hilbert space embedding of distributions (KME)—which generalizes the feature map of individual points to probability measures—has emerged as a powerful machinery for probabilistic modeling, machine learning, and causal discovery. The idea behind this framework is to map distributions into a reproducing kernel Hilbert space (RKHS) endowed with a kernel k. It enables us to apply RKHS methods to probability measures and has given rise to a great deal of research and novel applications of kernel methods.

Given an i.i.d. sample $x_1, x_2, \ldots, x_n$ from $\mathbb{P}$, the most natural estimate of the embedding $\mu_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}[k(X,\cdot)]$ is an empirical average $\hat{\mu}_{\mathbb{P}} = (1/n)\sum_{i=1}^{n} k(x_i,\cdot)$. In [35, 370], we showed that this estimator is not optimal in a certain sense. Inspired by James-Stein estimator, we proposed the so-called kernel mean shrinkage estimators (KMSEs) which improves upon the standard estimator. A suitable explanation for the improvement is a bias-variance tradeoff: the shrinkage estimator reduces variance substantially at the expense of a small bias. In addition, we presented a class of estimators called spectral shrinkage estimators in [395] which also incorporates the RKHS structure via the eigenspectrum of the empirical covariance operator. Our empirical studies suggest that the proposed estimators are very useful for "large p, small n" situations (e.g. medical data, gene expression analysis, and text documents).

A natural application of KME is in testing for similarities between samples from distributions. We refer to the distance between two distribution embeddings as the maximum mean discrepancy (MMD). We have formulated a two-sample test [142] (of whether two distributions are the same), and showed that the independence test (of whether two random variables observed together are statistically independent) is a special case. A further application of the MMD as independence criterion is in feature selection, where we maximize dependence between features and labels [143]. We have further developed alternative independence tests based on space partitioning approaches and classical divergence measures (such as the ℓ_1 distance and KL-divergence) [268]. Lastly, we also constructed the test for non-i.i.d. data such as time-series in [441].

Given that the MMD depends on the particular kernel that is chosen, we proposed two kernel selection strategies [494], the earlier one relying on a classification interpretation of the MMD, and the later one explicitly minimizing the probability of Type II error of the associated two-sample test (that is, the probability of wrongly accepting that two unlike distributions are the same, given samples from each).

We have also used the KME to develop a variant of an SVM which operates on distributions rather than points [478], permitting modeling of input uncertainties. One can prove a generalized representer theorem for this case, and in the special case of Gaussian input uncertainties and Gaussian kernel SVMs, it leads to a multiscale SVM, akin to an RBF network with variable widths, which is still trained by solving a quadratic optimization problem. In [356], we applied this framework to perform bivariate causal inference between X and Y as a classification problem on joint distribution $\mathbb{P}(X, Y)$. Another interesting application is in domain adaptation [407, 676]. This idea has also been extended to develop a variant of One-class SVM that operates on distributions, leading to applications in group anomaly detection [415].

A recent application uses kernel means in visualization. When using a power-of-cosine kernel for distributions on the projective sphere, the kernel mean can be represented as a symmetric tensor. In the context of diffusion MRI, this permits an efficient visual and quantitative analysis of the uncertainty in nerve fiber estimates, which can inform the choice of MR acquisition schemes and mathematical models [110, 388].

A natural question to consider is whether the MMD constitutes a metric on distributions, and is zero if and only if the distributions are the same. When this holds, the RKHS is said to be characteristic. We have determined necessary and sufficient conditions on translation invariant kernels for injectivity, for distributions on compact and non-compact subsets of $\mathbb{R}^d$ [253]: specifically, the Fourier transform of the kernel should be supported on all of $\mathbb{R}^d$. Gaussian, Laplace, and B-spline kernels satisfy this requirement. The MMD is a member of a larger class of metrics on distributions, known as the integral probability metrics (IPMs). In [16, 4], we provide estimates of IPMs on $\mathbb{R}^d$ which are taken over function classes that are not RKHSs, namely the Wasserstein distance (functions in the unit Lipschitz semi-norm ball) and the Dudley metric (functions in the unit bounded Lipschitz norm ball), and establish strong consistency of our estimators. Comparing the MMD and these two distances, the MMD converges fastest, and at a rate independent of the dimensionality d of the random variables – by contrast, rates for the classical Wasserstein and Dudley metrics worsen when d grows.

Embeddings of distributions can be generalized to yield embeddings of conditional distributions. The first application is to Bayesian inference on graphical models. We have developed two approaches: in the first [556, 602], the messages are conditional density functions, subject to smoothness constraints; these were orders of magnitude faster than competing nonparametric BP approaches, yet more accurate, on problems including depth reconstruction from 2-D images and robot orientation recovery. In the second approach [558], conditional distributions $P(Y|X = x)$ are represented directly as embeddings in the RKHS, allowing greater generality (for instance, one can define distributions over structured objects such as strings or graphs, for which probability densities may not exist). We showed the conditional mean embedding to be a solution to a vector valued regression problem [492], which allows us to formulate sparse estimates. The second application is to reinforcement learning. In [491], we estimate the optimal value function for a Markov decision process using conditional distribution embeddings, and the associated policy. This work was generalized to partially observable Markov decision processes in [462], where the kernel Bayes' rule was used to integrate over distributions of the hidden states.

Another important application of conditional mean embeddings is in testing for conditional independence (CI). We proposed a Kernel-based Conditional Independence test (KCI-test) [519] which avoids the classical drawbacks of CI testing. Most importantly, we further derived its asymptotic distribution under the null hypothesis, and provided ways to estimate such a distribution. Our method is computationally appealing and is less sensitive to the dimensionality of Z compared to other methods. This is the first time that the null distribution of the kernel-based statistic for CI testing has been derived. Recently, we proposed a new permutation-based CI test [390] that easily allows the incorporation of prior knowledge during the permutation step, has power competitive with state-of-the-art kernel CI tests, and accurately estimates the null distribution of the test statistic, even as the dimensionality of the conditioning variable grows.

Lastly, we have recently leveraged the KME in computing functionals of random variables $Z = f(X_1, X_2, \dots, X_n)$ [43], which is ubiquitous in various applications such as probabilistic programming. Our approach allows us to obtain the distribution embedding of Z directly from the embeddings of $X_1, X_2, \dots, X_n$ without resorting to density estimation. It is in principle applicable to all functional operations and data types, thank to the generality of kernel methods. Based on the proposed framework, we showed how it can be applied to non-parametric structural equation models, with an application to causal inference. As an aside, we have also developed algorithms based on distribution embedding for identifying confounders [422], which is one of the most fundamental problems in causal inference.

More information: https://ei.is.tuebingen.mpg.de/project/kernel-distribution-embeddings

Optimization and large scale learning

Optimization lies at the heart of most machine learning algorithms. The key aspects of the applications in which these algorithms are applied include: high-dimensional, noisy, and uncertain data; huge volumes of batch or streaming data; intractable models, low accuracy, and reliance on distributed computation or stochastic approximations. The success of most machine learning algorithms depends on how the optimization techniques can adapt and exploit these facets. Our interests are broadly divided into two categories, convex and non-convex methods.

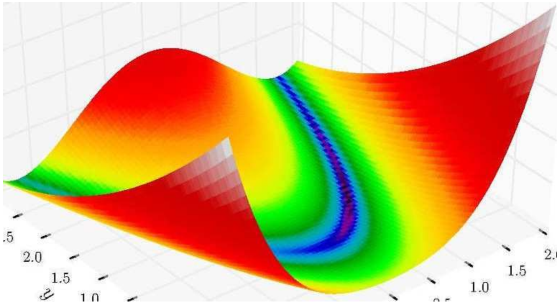


Figure 1.3: The Rosenbrock function, a non-convex function which serves as a test-bed for optimization algorithms (figure taken from Wikipedia)

Convex optimization In the realm of methods for convex optimization, we have addressed research challenges under various different problem settings. For large-scale problems, where scalability is an important aspect, a summary overview of large-scale aspects of convex optimization appears in our work [678]. A theoretically optimal large-scale convex method for problems with linear constraints is presented in [366] which develops a new stochastic alternating direction method of multipliers (ADMM) method that combines Nesterov's accelerated gradient methods with ADMM.

For learning classifiers in extremely large output spaces, we have proposed a parallelizable mixed-norm regularization approach leading to convex but non-smooth optimization in our recent work [316]. We show that the resulting models can be orders of magnitude smaller than most state-of-the-art methods and also lead to better generalization performance.

A contribution that lies at the interface of combinatorial and convex optimization is presented in [444]. General purpose convex methods often rely on key subroutines such as projection and proximity operators. Continuing our effort from the previous SAB assessment towards developing a library of highly tuned subroutines, e.g., for total variation, we extend to multivariate total variation in [365]. A flourishing sub-field of convex optimization is "Robust optimization" which seeks to optimize models under parameters/data uncertainty. It intersects with the usual min-max theory in statistics, and can be used to offer a different view of regularization in machine learning. In [30], we introduced the notion of robust optimization for matrix completion/recovery problems. The actual application was to recover correlation matrices under a simple bounded uncertainty model.

Non-convex Optimization In the domain of non-convex optimization for large-scale problems, our work [673] presents a simplified analysis of what, to our knowledge, is the first non-convex, non-smooth incremental proximal method. This work started in 2011; interestingly, in recent years, the interest in incremental methods has sky-rocketed, though the analysis is limited only to the convex case. Finally, we mention a new direction in nonconvex optimization offered by our recent work [443], which introduces "Geometric optimization" on the manifold of positive definite matrices. The underlying idea is to develop a theory of convexity along geodesics on the Positive Semi-Definite manifold. This work also identifies some basic calculus rules for detection and construction of geodesically convex functions on the Positive Definite manifold, and as an application presents new algorithms for solving maximum likelihood estimation for elliptically contoured distributions, which despite non-convexity remain tractable thanks to geodesic convexity.

More information: https://ei.is.tuebingen.mpg.de/project/optimization-and-large-scale-learning

Structure and dynamics of diffusion networks

We are interested in developing machine learning and large-scale data mining methods for the analysis, modeling and control of large real-world networks and processes that take place over them. We addressed mainly two problems: network inference and influence maximization.

Network inference Observing a diffusion process often reduces to noting when nodes (people, blogs, etc.) reproduce a piece of information, get infected by a virus, or buy a product. Epidemiologists can observe when people become ill but they cannot tell who infected them or how many exposures and how much time was necessary for the infection to take hold. In information propagation, we observe when a blog mentions a piece of information. However if, as is often the case, the bloggers do not link to their sources, we do not know where they acquired the information or how long it took them to post it. Finally, viral marketers can track when customers buy products or subscribe to services, but typically cannot observe who influenced customers' decisions, how long they took to make up their minds, or when they passed recommendations on to other customers. We observe where and when but not how or why information (be it in the form of a virus, a meme, or a decision) propagates through a population of individuals. The mechanism underlying the process is hidden. However, the mechanism is of outstanding interest, since understanding diffusion is necessary for stopping infections, predicting meme propagation, or maximizing sales of a product.

In our work, we formulated a generative probabilistic model of diffusion that aims to describe realistically how diffusion occurs over time in a network. First, we developed two algorithms, NetInf [133, 566] and MultiTree [457], which use submodular optimization to infer the structure of a diffusion network from diffusion traces. We demonstrate the effectiveness of our algorithm by tracing information cascades in a set of 170 million blogs and news articles from 3.3 million sites over a one year period. We find that the diffusion network of news tends to have a core-periphery structure with a small set of core media sites that diffuse information to the rest of the Web.

However, both algorithms force the transmission rates between all nodes to be fixed and static – and not inferred. To overcome this limitation, we then developed the algorithm NetRate [69, 375, 418, 512], which allows transmission at different rates across different edges, possibly time-varying, so that we can infer temporally heterogeneous interactions within a network. NetRate infers both structure and the temporal dynamics of the underlying network with provable guarantees using convex optimization. We study the evolution of information pathways in the online media space and find that information pathways for general recurrent topics are more stable across time than for on-going news events. Clusters of news media sites and blogs often emerge and vanish in matter of days for on-going news events. Major events involving civil population, such as the Libyan's Civil War, lead to an increased amount of information pathways among blogs as well as in the overall increase in the network centrality of blogs and social media sites.

Influence maximization In this problem, one aims to select the most influential source node set of a given size in a diffusion network. A diffusion process that starts in such an influential set of nodes is expected to reach the greatest number of nodes in the network. Although the problem depends dramatically on the underlying temporal dynamics of the network, this was largely unexplored. In our work, we developed an influence estimation, InfluMax [458], which accounts for the temporal dynamics underlying diffusion processes [512]. Later, we developed a very efficient influence estimation method, ContInEst [429], which can be used in combination with InfluMax to scale up influence maximization to networks with million of nodes.

More information: https://ei.is.tuebingen.mpg.de/project/structure-and-dynamics-of-diffusion-networks

Causal Inference

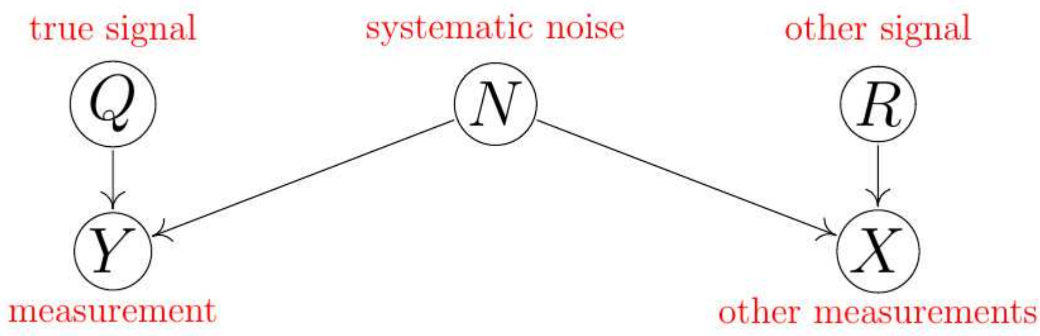


Figure 1.4: Visualization of a technique called "half-sibling regression" which corrects measurement errors. Measuring X allows us to remove parts of the systematic noise N from Y, which yields a better estimation of Q (the quantity of interest).

Causal Discovery In causal discovery, we try to learn a causal structure from data. This structure may be a directed acyclic graph (DAG) underlying a functional causal model (FCM), and possibly also the functions in the FCM. Without further assumptions, this goal is impossible to achieve: given only an observational distribution P, we can find an FCM generating P for any graph G s.t. P is Markovian w.r.t. G, i.e., the underlying graph is not identifiable. We investigate assumptions that make the graph identifiable from the observational distribution, and develop algorithms building on those assumptions. We now provide some examples, and refer to the literature for further papers that we cannot discuss below [353, 358, 401, 422, 518].

Example (1): Additive Noise Models. In additive noise models, the functional assignments used in the FCM are of the form $Z = f(X, Y) + N$. The subclass of linear functions and additive Gaussian noise does not lead to identifiability. This, however, constitutes an exceptional setting. If one assumes either (i) non-Gaussian noise, (ii) non-linear functions in the FCM [84, 624] or (iii) all noise variables to have the same variance, one can show that additive noise models are identifiable. Methods that are based on additive noise models perform above chance level not only on artifical data but also on the set of cause-effect pairs that we have collected over the last years [36]. A similar result holds if all variables are integer-valued [585] or if we interpret the additivity in $\mathbf{Z}/k\mathbf{Z}$ [200]. The concept of additive noise has been extended to time series, too [438], as well as to the identification of confounders [625].

Example (2): Information Geometric Causal Inference (IGCI). While the above methods inherently rely on noisy causal relations, statistical asymmetries between cause and effect also appear for deterministic relations. We have considered the case where $Y = f(X)$ and $X = f^{-1}(Y)$, for some invertible function f, the task being to tell which variable is the cause. The general assumption of independence of causal mechanisms [234] implies that $P(X)$ and $P(Y|X)$ should be independent if X causes Y. Choosing $P(X)$ and f independently implies that $P(Y)$ tends to have high probability density in regions where f^{-1} has large Jacobian. This observation can be made precise within an information theoretic framework [134, 562]. Applying a non-linear f to $P(X)$ decreases entropy and increases the relative entropy distance to Gaussians, provided that a certain independence between f and $P(X)$ is postulated which can be phrased as orthogonality in information space.

A second approach, for linear invertible relations between multi-dimensional variables, is related in spirit: if the covariance matrix of X and the structure matrix relating X and Y are chosen independently, directions with high covariance of Y tend to coincide with directions corresponding to small eigenvalues of A^{-1}, which can be checked by a formula relating traces of covariance matrices with traces of the product of structure matrices with their transpose [518, 571].

Example (3): Invariant Prediction. In many situations, we are interested in the system's behavior under a change of environment. Here, causal models become important because they are often invariant under those changes [470]. Following the assumption of independence of causal mechanisms, localized changes of some noises or mechanisms of a causal model will often leave other conditionals invariant, and thus a causal prediction (which uses only direct causes of the target variable as predictors) may remain valid even if we intervene on predictor variables or change the experimental setting. We can exploit this for causal discovery: given data from different experimental settings, we use invariance as a criterion to estimate the set of causal predictors for a given target variable. This method also leads to valid confidence intervals for causal relations [20].

Example (4): Hidden Confounding in Time Series. Assume we are given a multivariate time series $X_1, \ldots, X_L$ of measurements. In this project, our goal is to infer the causal structure underlying $X_1, \ldots, X_L$, in spite of a potential unobserved confounder $(Z_t)_{t\in\mathbb{Z}}$ (which other approaches such as Granger causality cannot handle). We assume a vector autoregressive causal model

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Restricting the model class to non-Gaussian independent noise $(N_t)_{t\in\mathbb{Z}}$ makes B and C essentially identifiable [351]. We show that $D = 0$ is another sufficient restriction of the model class.

Example (5): Causal Strength. In real-world applications a measure of the *strength* of a causal influence is often required. This is a challenging question, even if the causal directed acyclic graph and the joint distribution are perfectly known. We have formulated a set of postulates that a measure of causal strength of an arrow (or of a set of arrows) in a causal network should satisfy [94]. We show that none of the measures in the literature satisfies all postulates, and describe examples where they therefore lead to poor results. This includes well-known approaches like Granger causality and transfer entropy for time series, and also measures for general graphs. The main problem is to quantify which part of the statistical dependences between two variables is due to a direct causal influence of one on the other, and which part is due to other causal paths. We propose an information-theoretic measure that satisfies all our postulates. It coincides with mutual information for a DAG with two nodes; for more complex DAGs it correctly removes the information propagated via alternative paths.

Causal Inference in Machine Learning We believe that causal knowledge is not only useful for predicting the effect of interventions, but that in some scenarios causal ideas can also improve the performance of classical machine learning methods. Again, we concentrate only on two examples and refer to some other papers [90, 470].

Example (6): Semi-supervised Learning. Our work [470] discusses several implications of the independence of cause and mechanism (as a special case of a more general independence principle) for standard machine learning. Let us assume that Y is predicted from X. We have argued that semi-supervised learning (SSL) does not help if X is the cause of Y (“causal learning”), whereas it often helps if Y is the cause of X (“anticausal learning”). This is because additional observations of X only tell us more about $P(X)$ – which is irrelevant in the case of causal prediction because the prediction requires information about the *independent* object $P(Y|X)$. Our meta-study analyzing results reported in the SSL-literature supports this hypothesis: all cases where SSL helped where anticausal, confounded, or examples where the causal structure was unclear. To elaborate on the link between causal direction and performance of SSL, we studied the toy problem of interpolating a monotonically increasing function for the case where the relation between X and Y is deterministic [24]. In such a scenario $P(X)$ can be shown to be beneficial for predicting Y from X whenever $P(X|Y)$ and $P(Y)$ satisfy a certain independence condition which coincides with the one postulated in our work on information-geometric causal inference [134].

Figure 1.4, finally, visualizes the idea of a recent method for correcting measurement errors called “half-sibling regression:” subtracting the conditional expectation of Y given X from Y can provide a better estimation of the quantity Q of interest than the noisy measurement Y, which has been used in [349] to process astronomical light curves for the detection of exoplanets.

More information: https://ei.is.tuebingen.mpg.de/project/causal-inference

Probabilistic control

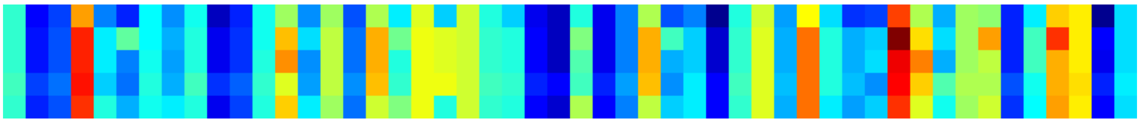


Figure 1.5: Systems of high variability can benefit from probabilistic dual control. To illustrate this, 50 different parameter sets (horizontal) are simulated for 5 different controllers (vertical). The colour indicates overall cost, from dark blue (low) to dark red (high).

Uncertainties have long been recognized as a key difficulty for control, deteriorating performance or even putting system safety at risk. This issue has been classically addressed by robust controller design, making use of a deterministic bound on the uncertainty and designing the controller for all possible uncertainty realizations. Tight uncertainty bounds are, however, difficult to obtain, limiting performance in practice, and robust techniques generally suffer from significant computational complexity.

The scope of this research is to develop new methods and tools for high performance and computationally efficient control of uncertain complex systems. We take a probabilistic approach, considering that the uncertainty affecting the system has a probabilistic nature. In contrast to common stochastic control methods, a key feature of the developed techniques is that they enable an online identification of the uncertainty distribution to maximize system performance, while providing (probabilistic) guarantees on the closed-loop system characteristics. The methods thereby leverage an interplay of probability, robustness, and adaption.

While uncertainties can enter controllers in various forms, we focus on model-based control techniques with model uncertainties due to high system complexity or due to the environment. The challenge is that learning happens in closed loop, i.e., 1) the resulting model is directly critical for both performance and safety at every time step and, 2) control and system identification have to be performed simultaneously.

Learning-based model predictive control Model predictive control (MPC) makes use of predictions of the system and the environment in order to optimally choose a sequence of control inputs that provides safety in the form of stability and constraint satisfaction and minimizes a performance objective.

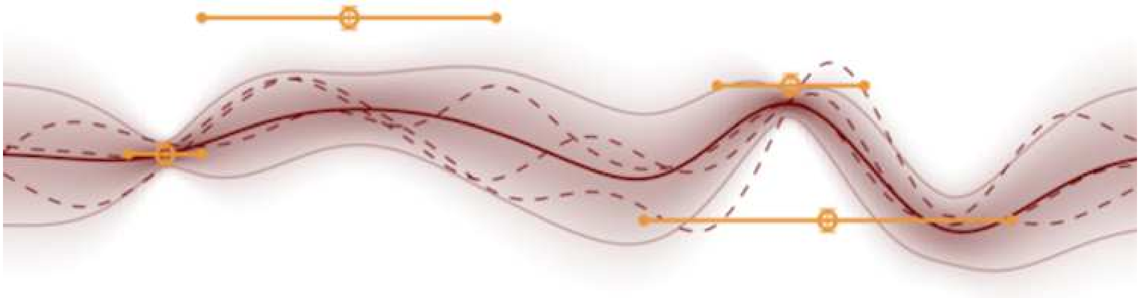
We investigate the use of learning-based identification methods for integration in an MPC controller, offering an ideal framework to incorporate probabilistic predictions based on online data. Due to their flexibility, Gaussian processes (GP) are used to model complex system behaviors and, most importantly, characterize the corresponding model uncertainty. The work [16] proposed a modeling and control approach using Gaussian processes to predict time-periodic errors that are then compensated by a predictive controller, which was demonstrated for high precision control of a telescope mount.

Dual control A key challenge of simultaneous control and system identification is that the control action influences not only the performance, but also the uncertainty in the dynamics. This is also called the exploration-exploitation trade-off. Dual control offers one way to address this trade-off by considering the control of the dynamical system augmented by the unknown parameters. While the exact dual controller is intractable, available approximations generally cannot maintain all core features: caution, exploration and the future value of information.

We have derived a tractable approximation of the dual control formulation that aims at maintaining these features, in particular the value of information, and have extended the technique to general nonlinear systems. The developed dual control framework has been applied in simulation to the problem of building control, providing simultaneous identification and control guided by electricity prices, weather, and occupancy.

More information: https://ei.is.tuebingen.mpg.de/project/probabilistic-control

Probabilistic numerics



Artificial intelligent systems build models of their environment from observations, and choose actions that they predict will have beneficial effect on the environment's state. The mathematical models used in this process call for computations that have no closed analytic solution. Learning machines thus rely on a whole toolbox of numerical methods: high-dimensional *integration* routines are used for marginalization and conditioning in probabilistic models. Fitting of parameters poses nonlinear (often non-convex) *optimization* problems. Predicting dynamic changes in the environment involves solving *differential equations*. In addition, there are special cases for each of these tasks in which the computation amounts to large-scale *linear algebra* (i.e. Gaussian conditioning, least-squares optimization, linear differential equations). Traditionally, machine learning researchers have served these needs by taking numerical methods "off the shelf" and treating them as black boxes.

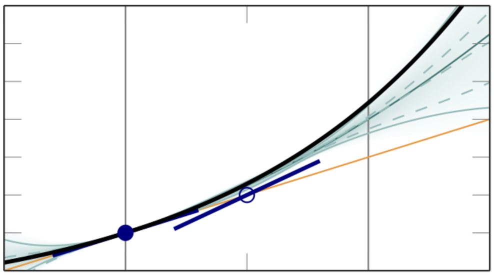
Since the 1970s, researchers like Wahba, Diaconis, and O'Hagan repeatedly pointed out that, in fact, numerical methods can themselves be interpreted as statistical rules—more precisely, as acting machines, since they take decisions about which computations to perform: they estimate an unknown intractable quantity given known, tractable quantities. For example, an integration method estimates the value of an integral given evaluations of the integrand. This is an abstract observation, but Diaconis and O'Hagan separately made a precise connection between inference and computation in the case of integration: several classic quadrature rules, e.g. the trapezoid rule, can be interpreted as the maximum a posteriori (MAP) estimator arising from a family of Gaussian process priors on the integrand.

Over recent years, the research group on probabilistic numerics has been able to add more such bridges between computation and inference across the domains of numerical computation, by showing that various basic numerical methods are MAP estimates under equally basic probabilistic priors: quasi-Newton methods, such as the BFGS rule, arise as the mean of a Gaussian distribution over the elements of the inverse Hessian matrix of an optimization objective [91, 406, 448]. This result can be extended to linear solvers [49], in particular the linear method of conjugate gradients (Gaussian regression on the elements of the inverse of a symmetric matrix). Regarding ordinary differential equations, some Runge-Kutta methods can be interpreted as autoregressive filters [397], returning a Gaussian process posterior over the solution of a differential equation.

The picture emerging from these connections is a mathematically precise description of computation as the active collection of information. In this view, the analytic description of a numerical task provides a prior probability measure over possible solutions, which can be concentrated through conditioning on the result of tractable computations. Many concepts and philosophical problems from statistics carry over to computation quite naturally, with two notable differences: first, in numerical "inference" tasks, the validity of the prior can be analyzed to a higher formal degree than in inference from physical data sources, because the task is specified in a formal (programming) language. Secondly, since numerical routines are the bottom, "inner loop" layer of artificial intelligence, they must curtail computational complexity. This translates into a constraint on acceptable probabilistic models—most basic numerical methods make Gaussian assumptions.

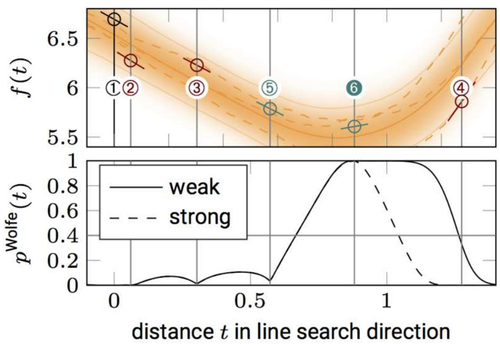
In the machine learning context, the description of computation as the collection of information has opened a number of research directions:

(1) Once it is clear that a numerical method uses an implicit prior, it is natural to adapt this prior to reflect available knowledge about the integrand. This design of "customized numerics" was used in a collaboration with colleagues at Oxford to build an efficient active integration method that outperforms Monte Carlo integration methods in wall-clock time on problems of moderate dimensionality [396].

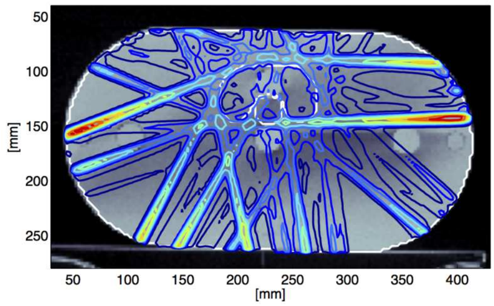


(2) Many numerical problems are defined relative to a setup that is itself uncertain to begin with. Once numerical methods are defined as probabilistic inference, such uncertainty can often be captured quite naturally. In a collaboration with colleagues in Copenhagen, it was shown [354, 369, 399] how uncertainty arising from a medical imaging process can be propagated in an approximate inference fashion to more completely model uncertainty over neural pathways in the human brain.

(3) Explicit representations of uncertainty can also be used to increase robustness of a computation itself. Addressing a pertinent issue in deep learning, we constructed a line search method [321]—a building block of nonlinear optimization methods—that is able to use gradient evaluations corrupted by noise. The resulting method automatically adapts step sizes for stochastic gradient descent.



(4) More generally, it is possible to define *probabilistic numerical methods*: Algorithms that accept probability measures over a numerical problem as inputs, and return another probability measure over the solution of the problem, which reflects both the effect of the input uncertainty, and uncertainty arising from the finite precision of the internal computation itself. A position paper [37] motivates this class of algorithms, and suggests their use for the control of computational effort across composite chains of computations, such as those that make up intelligent machines. In collaboration with the Optimization group at the German Cancer Research Center we developed approximations to propagate physical uncertainties through the optimization pipeline for radiation treatment, to lower the risk of complications for patients [111, 437].



In a separate but related development, a community has also arisen around the formulation of global optimization as inference, and the formulation of sample-efficient optimization methods. These *Bayesian Optimization* methods can, for example, be used to structurally optimize and automate the design of machine learning models themselves. We contributed to this area with the development of the Entropy Search [135] algorithm that automatically performs experiments expected to provide maximal information about the location of a function's extremum.

Probabilistic numerics is emerging as a new area at the intersection of mathematics, computer science and statistics. As co-founders, the research group on probabilistic numerics plays a central role in its development. The wider Intelligent Systems community, with their intractably large data streams and non-analytic model classes, are simultaneously contributors and beneficiaries: equipping computational routines with a meaningful notion of uncertainty stands to increase both the efficiency and reliability of intelligent systems at large.

More information: https://ei.is.tuebingen.mpg.de/project/probabilistic-numerics

Computational photography

$$y = Z_y^{\mathsf{T}} \sum_{r=0}^{p-1} C_r^{\mathsf{T}} F^{\mathsf{H}} \operatorname{Diag}(F Z_f\, f^{(r)}) F \operatorname{Diag}(w^{(r)}) C_r\, x$$

Figure 1.6: In many real-world imaging applications, the common assumption of stationary blur does not hold. Examples that exhibit non-stationary blur include e.g. camera shake, optical aberrations, and atmospheric turbulence. We derived a mathematically sound and physically well-motivated model, which allows to express and efficiently compute spatially-varying blur [570] . Our "Efficient Filter Flow" framework substantially broadens the application range of image deconvolution methods. In a number of challenging real-world applications [26, 320, 472, 535, 537, 570] we demonstrated both the validity and versatility of our approach.

Digital image restoration as a key area of signal and image processing aims at computationally enhancing the quality of images by undoing the adverse effects of image degradation such as noise and blur. It plays an important role in both scientific imaging and everyday photography. Using probabilistic generative models of the imaging process, our research aims to recover the most likely original image given a low-quality image, or image sequence. In the following we give a number of illustrative examples to highlight some of our work.

Spatially varying blurs: our work on blind deconvolution of astronomical image sequences [620] can recover sharp images through atmospheric turbulence, but is limited to relatively small patches of the sky, since the image defect is modeled as a space-invariant blur. Images that cover larger areas require a convolutional model allowing for space-variant blur. In [570] we proposed such a model based on a generalization of the short-time Fourier transform, called Efficient Filter Flow (EFF), which is illustrated in Fig. 1.6. Deconvolution based on EFF successfully recovers a sharp image from an image sequences distorted by air turbulence (see Fig. 1.7).

Super-resolution: we generalized our online method for the *multi-frame blind deconvolution* problem [620] to account for the adverse effect of saturation and to enable super-resolution given a sequence of degraded images [184]. By drawing a connection between Fraunhofer diffraction and kernel mean maps we are able to show that under certain imaging conditions imaging beyond the physical diffraction limit is in principle possible [423].

Removing camera shake: photographs taken with long exposure times are affected by camera shake creating smoothly varying blur. Real handheld camera movement involves both translation and rotation, which can be modeled with our EFF framework for space-variant blur [597]. We were also able to recover a sharp image from a single distorted image, using sparsity-inducing image priors and an alternating update algorithm. The algorithm can be made more robust by restricting the EFF to blurs consistent with physical camera shake [537]. This leads to higher image quality (Fig. 1.8) and computational advantages. To foster and simplify comparisons between different algorithms removing camera shake, we created a public benchmark dataset and a comparison of current methods [483].

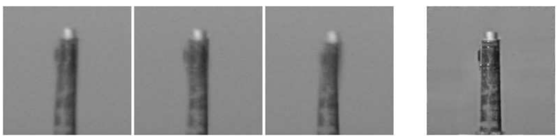


Figure 1.7: Our Efficient Filter Flow model [570] combined with our online multi-frame blind deconvolution [620] allows to restore a single sharp image given a sequence of blurry images degraded by atmospheric turbulence.

Figure 1.8: Example result of our proposed method for fast removal of non-uniform camera shake [537]. The left image shows the original image blurred due to camera shake during exposure, the right image shows the result of our proposed method.

Correcting lens aberration: even good lenses exhibit optical aberration when used with wide apertures. Similar to camera shake, this creates a certain type of blur that can be modeled with our EFF framework; but optical aberrations affects each color channel differently (chromatic aberration). We measured these effects with a robotic setup and corrected them using a non-blind deconvolution based on the EFF framework [535]. We were also able to implement a blind method rectifying optical aberrations in single images [472]. An example is shown in Fig. 1.9. The key was to constrain the EFF framework to rotationally symmetric blurs varying smoothly from the image center to the edges. We have recently been able to formulate this as non-parametric kernel regression, enabling faithful optical aberration estimation and correction [320], which might lead to new approaches in lens design.

Figure 1.9: To test the capability of our approach to optical aberration correction, we built a camera lens containing a single glass element only. An example image is shown on the left, which exhibits strong blur artifacts especially towards the corners of the image. In contrast, the restored image (right) is of high quality and demonstrates that our method [472] is able to correct for most degradations caused by optical aberrations from a single input image only.

Denoising: another classical image distortion is *noise*. Noise can exhibit different structure, e.g., additive white Gaussian, salt-and-pepper, JPEG-artifacts, and stripe noise, depending on the application.

In astronomical imaging, dim celestial objects require very long exposure times, which causes high sensor noise. It is common to subtract a dark frame from the image — an image taken with covered lens, containing only sensor noise. The difficulty with this is that the sensor noise is stochastic, and image information is not taken into account. We studied the distribution of sensor noise generated by a specific camera sensor and proposed a parameterized model [538]. Combined with a simple image model for astronomical images, this gives superior denoising.

Multi-scale denoising: noise usually has a more damaging effect on the higher spatial frequencies, so most algorithms focus on those. However, if the noise variance is large, also lower frequencies are distorted. We were able to significantly improve the performance of many methods for that setting by defining a multi-scale meta-procedure, leading to a DAGM 2011 prize [539].

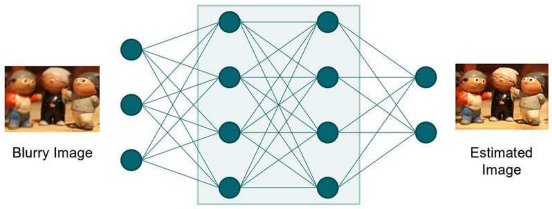


Figure 1.10: Schematic illustration of casting image restoration as a learning problem. A neural network (middle) is trained with degraded/clean image pairs to learn a mapping from a degraded image (left) to a restored image with enhanced quality (right).

Image restoration as a learning problem: since denoising can be also seen as a non-trivial mapping from noisy to clean images as sketched in Fig. 1.10, we were particularly interested whether a learning-based approach can be applied. Using very large data sets, we trained a multi-layer perceptron (MLP) that is able to denoise images better than all existing methods, leading to the new state-of-the-art denoising method [450].

Such a discriminative approach turned out to be effective also for other image restoration tasks such as inpainting [402], non-blind deconvolution [424], as well as blind image deconvolution [26, 320].

More information: https://ei.is.tuebingen.mpg.de/project/computational-imaging

Medical and neuroscientific imaging

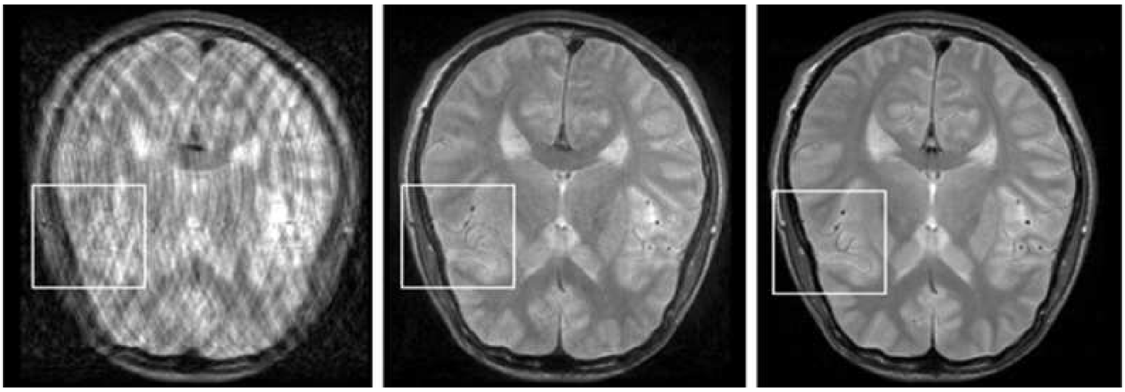


Figure 1.11: Retrospective, fully data-driven removal of motion-related artefacts in 2D RARE images. Left: motion corrupted image. Middle: reconstruction. Right: no-motion image.

Modern medical imaging is a blend of 20th century breakthroughs in physics and information processing. Although there is an ongoing advancement of technology, many of the problems remain unsolved. Using machine learning and image processing techniques for more efficient and intelligent use of acquired data seems a promising solution to some of the obstacles. For example, understanding the statistical properties of medical images allows making certain prior assumptions on unobserved or corrupted aspects of the data, which makes it possible to improve the image quality and aid the medical diagnostics.

PET-MR systems combine functional information from Positron Emission Tomography (PET) with structural information from Magnetic Resonance (MR) Imaging. They demand PET attenuation (or μ) map to be determined from MR data, which constitutes an ill-posed problem. We have developed a Gaussian process prediction framework, trained on co-registered MR and μ-maps, and evaluated it on brain and whole body data. A collaborative effort with the university hospital, Tübingen, the method has been patented and licensed to Siemens AG. Further research focused on evaluating the influence of positioning aids [191], improving the robustness of the method in presence of MR image artifacts [106] and the assessment of PET quantification accuracy in a pediatric patient collective [28].

Patient motion during long MR scans leads to severe non-local degradations and can render images unacceptable for medical diagnosis. We developed a retrospective (post-processing) approach called GradMC, which operates on the motion-corrupted images [98]. It automatically corrects for motion artifacts by searching for a motion-generating function whose inverse application optimizes a quality measure based on image statistics. Our GPU implementation processes a full 3D volume in a few minutes, which is acceptable in routine clinical use. In a follow-up work, we have come up with a fully retrospective non-rigid motion correction scheme that only needs raw data as an input [50].

Diffusion MRI (dMRI) allows us to investigate brain tissue microstructure non-invasively. Recently, supervised classification has been used to recognize patterns in this rich and complex data that allow us to automatically detect neurological diseases. We have contributed to this field by developing improved methods for feature extraction [345], by applying machine learning to specific diseases [56], and via a survey that summarizes the state of the art.

More information: https://ei.is.tuebingen.mpg.de/project/machine-learning-for-medical-and-neuroscientific-imaging

Inference in Astronomy

Figure 1.12: *From left to right:* (1) deepest existing image of NGC 5907, showing a recently discovered faint stellar stream (Martínez-Delgado et al., The Astronomical Journal., 689:184-193, 2008); (2-5) example images obtained by a Web search for NGC 5907 (manually removing images that show the stellar stream); (6) result obtained by our algorithm (tone-mapped to the histogram of the deep image).

Astronomy offers some of the most exciting opportunities for machine learning applications in the natural sciences. Satellite missions as well as ground-based observations have generated datasets of unprecedented quality and size. New subfields of astronomy such as exoplanet science have emerged, impossible without large scale data analysis. However, by and large, data analysis in astronomy is still rather conservative, and there are thus large opportunities for data-driven modeling and probabilistic inference. Our work in this field benefits from methods developed within several projects in the department, and some of it is thus discussed elsewhere (see page 23 where an astronomical image deconvolution method [184] is discussed).

Combining hetergeneous imaging The Web contains millions of astronomical images that contain scientifically valuable information about the night sky. However, these images are difficult to use for scientific purposes because images rendered for human visual consumption are usually processed with non-linear tone mappings. We developed a system that builds a high dynamic-range and wide-angle image of the night sky by combining a large set of such images [361]. It computes a "consensus" image whose pixel brightness *rank* information tries to agree with the individual images. The complexity of the algorithm is linear in the number of images. It permits discovery of astronomical objects or features that are not visible in any of the input images taken individually, as shown by a faint tidal stellar stream (see Figure 1.12). More importantly, however, it permits scientific exploitation of a huge source of astronomical images that would not be available to astronomical research without our automatic system The system was featured on the BBC Stargazing Live program.

Half-sibling regression for modeling systematic errors
The Kepler space observatory, launched in 2009, observes a tiny fraction of the Milky Way in search for exoplanets, monitoring the brightness of 150,000 stars. Some of them exhibit periodic descreases of brightness in their light curves due to partial occlusions caused by exoplanets. These measurements, however, are corrupted with systematic noise due to the telescope. However, since the stars can be assumed to be causally independent of each other (they are light years apart) as well as of the instrument noise, it turns out that we can denoise the signal of a single star by removing all information that can be explained by the measurements of the other stars. Under the assumption that the systematic noise acts in an additive manner, we provide strong theoretical guarantees regarding the quality of the reconstruction [349].

Exoplanet search & characterization We applied the half-sibling regression model to perform the first systematic search for transiting exoplanet signals in new data from NASA's K2 Mission. Using this method, we discovered 36 new planet candidates—of which 21 were subsequently validated as bona fide exoplanets (Montet et al., Astrophysical Journal, 809, 25, 2015)—and demonstrated that these data could be used for planet detection despite the large systematic errors [38]. We used principal component analysis to reduce the dimensionality of the predictors and extended the half-sibling methodology to jointly fit for the regression weights and the exoplanet properties. This increased the sensitivity of our search to small signals and allowed the propagation of uncertainties in the systematics model to our estimates of the physical parameters.

More information: https://ei.is.tuebingen.mpg.de/project/astronomy

Robot skill learning

Figure 1.13: The robot learning lab aims at finding a general approach to acquiring motor skills. Developed algorithms and methods are tested in skill learning tasks such as playing table tennis.

Creating autonomous robots that can learn to assist humans in daily life situations is a fascinating challenge for machine learning. We focus on the first step of creating robots that can learn to accomplish many different tasks triggered by an environmental context or a higher-level instruction and plan to obtain a general approach to motor skill learning. We focus on (1) domain-appropriate machine learning approaches that allow for better control, imitation of behavior and self-improvement, as well as (2) new robotics approaches to create more appropriate systems for high-speed skill learning.

Starting from theoretically sound robotic control structures for task representation and execution, we replace analytic modules with more flexible learned ones [468]. To this end, we tackle problems such as accurate but compliant execution in joint-space [583] or task-space [150], learning of elementary behaviors using combination of imitation and reinforcement learning [178, 238], hierarchical composition of behaviors, and parsing complex demonstrations into elementary behaviors.

Mimicking how children learn new motor tasks, we have used imitation to initialize to learn libraries of elementary primitives, and subsequently reinforcement learning to improve the performance. We have learned elementary tasks such as Ball-in-a-Cup or bouncing a ball [178, 238] and gradually moved to more complex ones. As the benchmark of complex behavior, we chose the task of returning table tennis balls over the net. We created a parser that segments movements of a human teacher into elementary movements [594, 596]. These then train the single elementary movements [178, 238]. Novel behaviors, modulated by the opponent's incoming ball, are composed by mixing motor primitives [465, 581]. With the help of this method, the robot table tennis player learnt with this method successfully returns of 97% of the balls played against a ball gun. Current approaches are likely to achieve even better results [347] and use accurate prediction of the human opponent's behavior before the opponent even touched the ball [459].

More information: https://ei.is.tuebingen.mpg.de/project/robot-skill-learning

Reinforcement learning

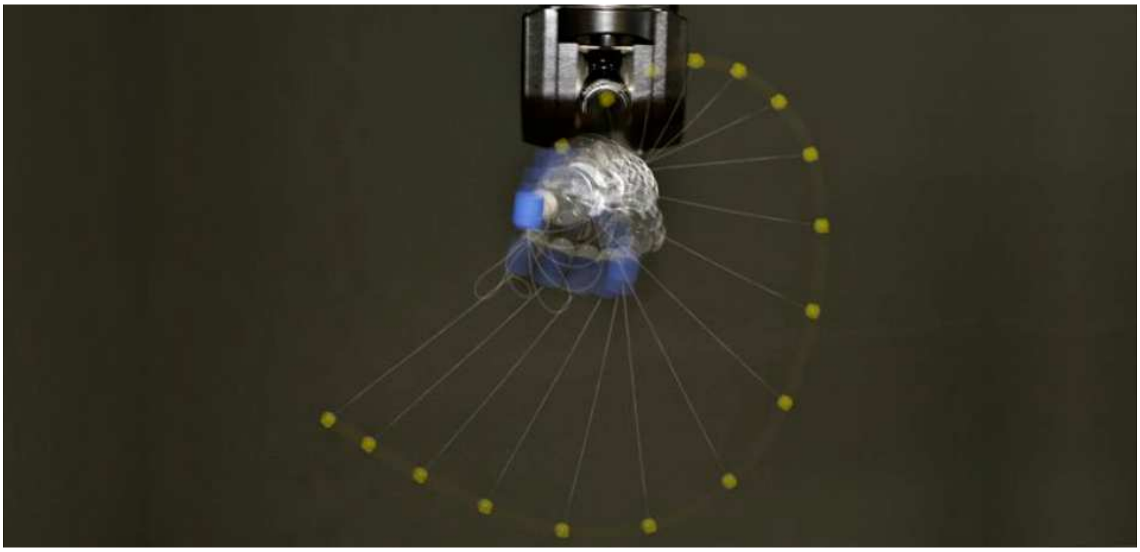


Figure 1.14: Ball-in-a-cup was learned with the EM-like policy search approach called PoWER [178].

Reinforcement learning ranks among the biggest challenges for machine learning. Just controlling a known dynamical system is hard on its own – interacting with an unknown system poses even harder decision problems, such as the infamous exploration-exploitation tradeoff. Most research in this area is still confined to theoretical analysis and simplistic experiments, but the promise of autonomous machines justifies the effort. Over the past years, members of the department contributed to reinforcement learning in theory and experiment.

Non-Parametric Dynamic Programming [530] showed that a non-parametric kernel density representation of system dynamics unifies several popular policy evaluation methods: their Galerkin method joins Least-Squares Temporal Difference learning, Kernelized Temporal Difference learning, and a type of discrete-state Dynamic Programming, as well as a novel method of improved performance.

EM-like Reinforcement Learning Policy search, a successful approach to reinforcement learning, directly maximizes the expected return of a policy – in contrast to value function approximation, which derives policies from a learnt value function. However, few of its variants scale to many dimensions, as they are based on gradient descent over many trials. To improve efficiency, [178] reduced the problem to reward-weighted imitation, treating rewards received after actions as improper probabilities indicating the actions' success. Their idea resembles Expectation Maximization, giving good actions a higher probability to be re-used. This framework also unifies previous algorithms, and allows the derivation of novel ones, such as episodic reward-weighted regression and PoWER.

Relative Entropy Policy Search Policy improvements in policy search often invalidate previously collected information, causing premature convergence and implausible solutions. These problems may be addressed by constraining the information loss. Relative Entropy Policy Search (REPS) bounds the information loss while maximizing expected return [586]. REPS differs significantly from previous policy gradient approaches. It yields an exact update shown to work well on reinforcement learning benchmarks. REPS can be generalized hierarchically [455] using a gating network to choose among several option policies. This hierarchical REPS learns versatile solutions while increasing learning speed and the quality of the learnt policy.

Bayesian reinforcement learning Probability theory gives a uniquely coherent answer to the exploration-exploitation dilemma: From the Bayesian perspective, reinforcement learning is about including possible future observations in considerations about optimal behavior. Since probabilistic models can predict future data, this process can be rigorously formalized. It amounts to modeling knowledge as an additional dynamic variable to be controlled. In general, the combinatorial number of possible futures is intractable; however, [540] showed that the Gaussian process (GP) framework, in which predictions involve linear algebra calculations, allows approximating optimal exploration-exploitation with classic numerical methods for the solution of stochastic differential equations.

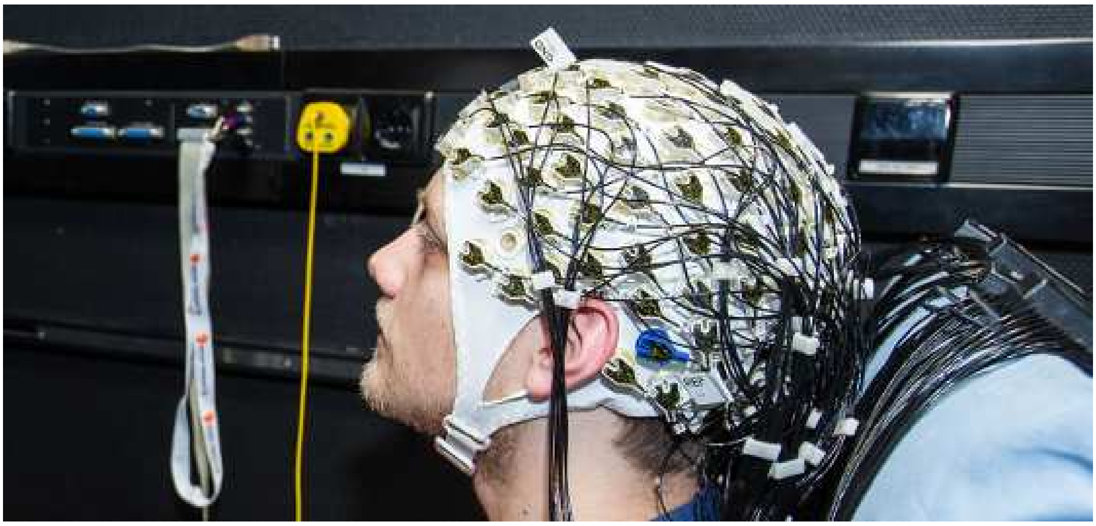
Reinforcement learning with Gaussian Processes: [279] used GPs for approximate dynamic programming in reinforcement learning, as probabilistic function approximators for the value function, and as models of the system dynamics. Using the predictive uncertainty for guidance, active learning methods could explore the state space efficiently. [555] proposed a particularly efficient use of GPs for optimal control over continuous states for non-bifurcating systems with low sampling rate. In their work, GPs capture information gained, as well as remaining uncertainty due to noise and lack of experience. The system's behavior is predicted by propagating state and action distributions through time, tractability is achieved approximating distributions by moment matched Gaussians. This "virtual simulation" is used to optimize the control policy. Their algorithms learn from even limited interactions with the environment due to the power of using probabilistic forward models for such indirect experience rehearsal.

Apprenticeship learning via inverse reinforcement learning Unguided exploration can be hazardous for systems like robots. This issue is addressed by imitation learning from example actions provided by an expert, where the autonomous agent learns a policy generalizing the demonstrations to new states. This behavioral cloning may fail when the dynamics of expert and learner differ. Indeed, even simple repetition of the expert's actions does not always yield the same results. An alternative is to infer the expert's reward function from the expert's behavior, then use it to learn in the new system. This avoids exhaustive exploration by searching for policies close to the expert's. Previous work required a model of the expert's dynamics, but [508] presented a model-free inverse reinforcement learning algorithm, using importance sampling to adapt expert examples to the learner's dynamics. Tested on several benchmarks, the algorithm proved more efficient than the state of the art. Generalization in both forward and inverse reinforcement learning depends on the projection of states onto features to describe reward and value function. Features, especially visual ones, are often subject to noise, for example in robot grasping and manipulation tasks. To solve this problem, [463] combined control and structured output prediction over Markov Random Fields to represent the action distribution. Their method is robust to noise in a grasping task, and can also be used in other applications requiring control from vision.

Data-dependent Analysis of Reinforcement Learning Many analyses of reinforcement learning focus on worst-case scenarios, although reality is often not adversarial. [536] used PAC-Bayesian inequalities for martingales in a data-dependent analysis of the exploration-exploitation trade-off [138]. We studied stochastic multi-armed bandits with side information (also known as contextual bandits), a general framework where at each round of the game the agent is presented with side information (e.g., symptoms of a patient in a medical application) and has to find the best action (e.g., the best drug to prescribe given the symptoms). This model class is also used for personalized advertising on the internet. Our analysis includes the actual usage of side information by the algorithm, rather than the total amount of side information provided. This allows offering a lot of side information and letting the algorithm decide what is relevant, improving the run time of the algorithm exponentially over the state of the art.

More information: https://ei.is.tuebingen.mpg.de/project/reinforcement-learning

Brain-computer interfacing



Brain-computer interfaces (BCIs) translate neural recordings into signals that may be used for communication and/or the control of neuroprosthetic devices. Research in this domain poses interesting challenges to machine learning, because data is typically scarce, noisy, and non-stationary [186]. Furthermore, good decoding algorithms are contingent on domain knowledge that is not readily available and difficult to incorporate into traditional statistical methods. Accordingly, we employ machine-learning methods to study neural processes involved in BCI-control and use these insights to develop novel decoding algorithms and enhance experimental paradigms.

Machine learning algorithms for brain-state decoding A crucial aspect of our work is the development of algorithms for real-time brain-state decoding. We have developed a graphical model decoding framework for ERP-based visual speller systems [187]. Furthermore, we were the first to successfully apply the framework of multi-task learning to the domain of BCIs [560]. As the signal characteristics used by subjects to control a BCI share common aspects, the incorporation of data from previously recorded subjects substantially decreases calibration time and enhances overall decoding performance [13]. Besides working on methods for real-time decoding, we also develop tools to investigate the neural basis of disorders of cognition [14, 46]. This is essential to understand how the diseased brain differs from the one of healthy subjects, which has implications for the design of BCI systems for patient populations.

Brain-computer interfaces for communication Building upon our machine-learning methods, we have investigated the neural basis of the ability to operate a BCI in healthy subjects and in patient populations. We were able to show that the configuration of large-scale cortical networks, as represented in high-frequency gamma-oscillations of the brain's electromagnetic field, influences a subject's ability to communicate with a BCI [149, 174]. Building upon these insights, we have developed a novel class of BCIs for patients in late stages of amyotrophic lateral sclerosis [81, 346].

Brain-computer interfaces for rehabilitation While BCIs were initially conceived as communication devices for the severely disabled, we have argued that they can also be used for stroke rehabilitation [183]. By combining a BCI with a seven degrees-of-freedom robotic arm, serving an exoskeleton, we developed a brain-controlled rehabilitation robot supporting patients with chronic stroke in self-regulation of sensorimotor brain rhythms [194]. The concept of brain-controlled rehabilitation robotics can be extended to systems that monitor patients' learning progress to adapt the rehabilitation exercise in real-time [65].

More information: https://ei.is.tuebingen.mpg.de/project/brain-computer-interfaces

Learning and inference for neuroscience

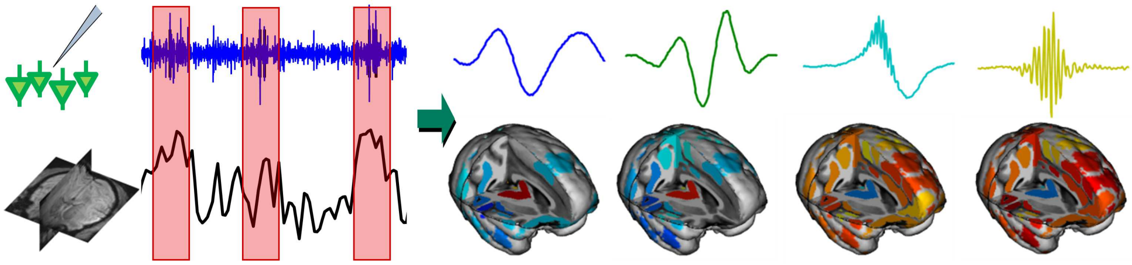


Figure 1.15: Left: example of LFP signal (blue) and concurrent fMRI activity (black), repetitive dynamical patterns appear in red blocks. Right: learned dictionary of dynamical patterns, together with their brain-wide signature.

Mammalian brains efficiently combine perception, decision making, and motor commands in hundreds of milliseconds. Machine learning can help understand these fascinating distributed information processing capabilities.

A first application of machine learning algorithms to neuroscience is the extraction of information from complex brain signals, focusing on their dynamical aspects. This includes the design of non-parametric statistical dependency measures for time series using an implicit mapping in a Reproducing Kernel Hilbert Space to capture complex non-linear dependencies between brain rhythms [441], as well as dictionary learning techniques that automatically identify the transient dynamical patterns in ongoing Local Field Potentials (LFP) of a given brain structure that have an impact on the activity of the whole brain (see Figure 1.15) [160].

A second objective is to infer causal statements about the organization of underlying neural mechanisms. One key application is the estimation of the direction and strength of information flow across brain networks. We first studied the communication between multiple locations of the visual cortex using an information theoretic measure of Granger causality, Transfer Entropy, computed between LFP signals during visual stimulation [224]. Our results suggest the presence of waves propagating along the cortical tissue along the direction of maximal flow of information, routing visual information across the visual cortex [34]. We also study novel causality principles proposed by our department as an alternative to the Granger causality framework widely used in neuroscience. In particular, we developed a new causal inference method for time series based on the postulate of Independence of Cause and Mechanism. We provided theoretical guarantees for this approach, which outperformed Granger causality in experimental LFP signals [353].

Causal inference techniques can also be used to assess which aspect of brain activity affects the behavioral outcome of an experiment. Causal terminology is often introduced in the interpretation of neuroimaging data without considering the empirical support for such statements. We investigated which causal statements are warranted and which ones are not supported by empirical evidence [46]. Our work provides the first comprehensive set of causal interpretation rules for neuroimaging results.

Beyond providing powerful data analysis techniques, machine learning can help understand brain function from a theoretical perspective. We investigated a standard model of neuronal adaptation – spike-timing dependent plasticity – in this way and showed it can be viewed as a stochastic gradient descent of a neuromodulatory reward function implementing a form of empirical risk minimization [95, 473]. These results help understand optimality properties of biological learning.

More information: https://ei.is.tuebingen.mpg.de/project/machine-learning-and-neuroscience

Psychophysics and computational models of behaviour

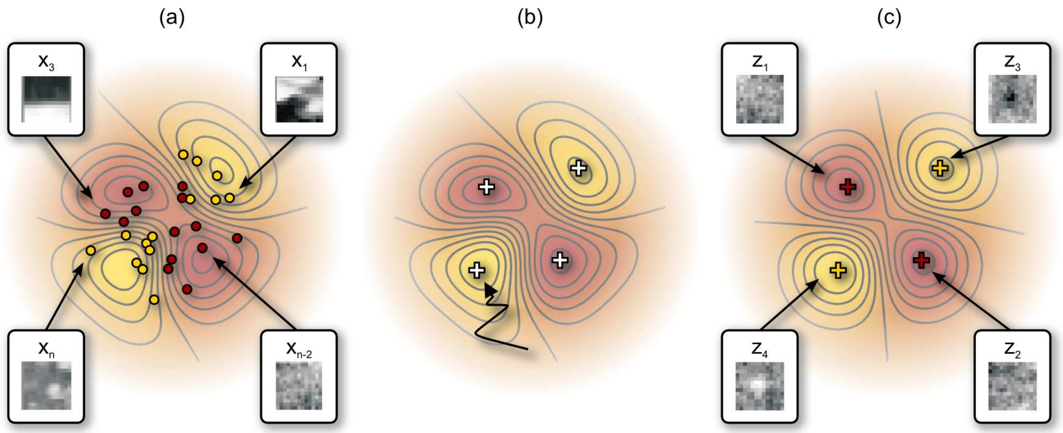


Figure 1.16: Visualization of the machine learning approach to uncover critical features. Consider the problem of predicting for an image patch whether it is likely to be a saccade target or not. Panel (a) depicts image patches that have (yellow) or have not (red) been saccade targets. The yellow-to-red gradient represent the response of a SVM with a Gaussian kernel trained on the patches. Panel (b): Gradient descent on the function that the SVM implements, extrema are indicated. Panel (c): Using the extrema, a network with four kernels centered on them approximates the SVM. (Adapted from [289].)

This project investigates human perception, combining psychophysical experiments and computational modeling. Currently we have four main research foci:

First, we develop methods to gain more information from psychophysical data. We characterized serial dependencies in behavioral responses, and introduced methods to correct for them [63]. We wrote a software package to perform Bayesian inference for the psychometric function for non-stationary data. In addition, we introduced the use of spatial point processes to characterize eye-movement fixation patterns [41, 99]. Finally, we continue to improve machine learning methods to uncover the critical features observers use in complex perceptual tasks [118, 148]) (c.f. [289] and Figure 1.16).

Our second focus is the development of an image-based model of spatial vision. We integrated the large psychophysical literature on simple detection and discrimination experiments and proposed a model based on maximum-likelihood decoding of a population of model neurons predicting several data sets simultaneously, using a single set of parameters [108].

Third, the project seeks to understand the fundamental questions of characterizing the computational principles underlying lightness perception, treating lightness perception as one example of how to rigorously study the visual mechanisms translating ambiguous retinal input into perceptually and psychologically relevant categories [22, 23, 117].

Finally, we are moving towards understanding perceptual causality. Recently there has been considerable progress in understanding causal inference by viewing it as a machine learning problem. This is relevant to perception since animals cannot operate based on the assumption of independent and identically distributed (iid) data but need to employ suitable inference methods that work under changing distributions in order to produce robust perception. We study causal models of perception in a simple binary classification setting where the perceiver needs to distinguish cause and effect, or forward and backward, a task we recently studied in a computer vision context [363]

More information: https://ei.is.tuebingen.mpg.de/project/psychophysics-and-computational-models-of-behaviour

1.3 Equipment

Computing Infrastructure

Figure 1.17: Some of the nodes of the computer cluster.

The desktop computing environment of the Department of Empirical Inference is based on Intel PCs, and currently uses the centrally managed operating systems Ubuntu Linux, Microsoft Windows, and Mac OS X. Silent PCs with no moving parts are used to provide the best possible work environment.

For data storage the department currently offers 80TB of online file system storage on two file servers, used for shared data as well as personal home directories. Backups are done through a tape library system, as well as a disk based snapshotting system for the most important data. Off-site storage from RZ Garching is offered for long-term archival of scientific data.

For numerical experiments the group has access to a high performance computing cluster, shared by the departments of the MPI for Intelligent Systems. The cluster nodes and their supporting infrastructure are maintained by a central scientific facility. The Empirical Inference department is using a 45% share of the central cluster as of now. The cluster comprises 32 rack-mounted multi-processor nodes based on the x86_64 architecture, running Ubuntu Linux. The nodes are connected via 10 GBit/s fiber-optic connections. In total, the system consists of 1536 Intel and AMD 64-bit CPU cores with 18.5 terabytes of memory, as well as 74 modern nVidia GPUs and 100 terabytes of fast distributed storage space. The most powerful machines have 64 CPU cores and 1 terabyte of memory.

Robot Learning Lab: High-Speed Robot Arms

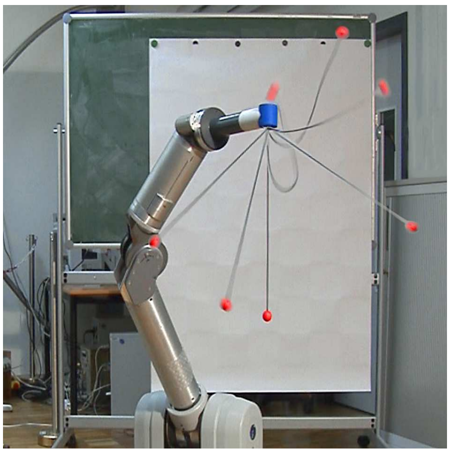


Figure 1.18: The Barrett WAM robot arm.

The Robot Learning Lab of the Empirical Inference Department focuses on finding task-appropriate machine learning methods for acquiring and refining motor skills. Current objectives include learning high-speed compliant control, learning simple skills (e.g., ball-in-a-cup or paddling balls), and larger composite tasks such as table tennis. These goals require a unique and maintenance-intensive set-up.

For fast as well as compliant control, we have a unique, custom-made high-voltage version of a Barrett WAM robot arm specifically designed for high-speed control while having seven degrees of freedom. Torque-level access to the robot, back drivability, and little backlash enables the use of this platform for learning control experiments and as a haptic input device during imitation learning or robot task assistance. The robot control makes use of a four camera high-speed (200 Hz) vision setup developed for that purpose. All software is based on a real-time Linux operating system and on the robot programming framework SL.

A tendon-driven pneumatic artificial muscle robot arm is being developed in order to study antagonistically actuated joints, allowing for lightweight segments, incorporating strong muscles, and using co-contraction for compliant control. We aim at approaching performance as observed in humans by designing such a system and enabling safer applications of learning control approaches whilst extending the variety of possible trajectories. We aim to use this new robot to show that learning can be particularly beneficial when classical methods fail.

Brain-Imaging Equipment

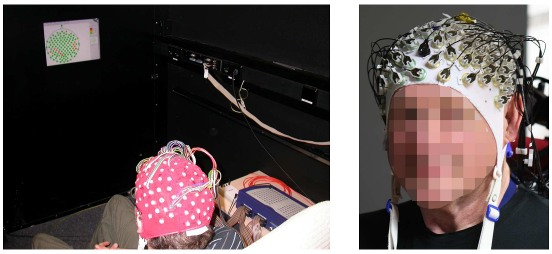


Figure 1.19: The BCI lab of the empirical inference department.

The QuickAmp 136 (Figure 1.19, left) is a high input-impedance amplifier capable of measuring and recording up to 128 channels of electroencephalographic (EEG) data, as well as eye (EOG), muscle (EMG), and other physiological signals. It is used for Brain-Computer Interface research, where brain signals with low noise are desirable, high temporal and spatial resolution, and where EOG and EMG signals are also required for artifact control. As of 2011, we also have a 128-channel active electrode system manufactured by BrainProducts GmbH, substantially reducing the preparation time of experimental studies and providing better signal-to-noise ratios outside of well-controlled laboratory environments (Figure 1.19, right). We have further acquired the Armeo Power, a robotic exoskeleton manufactured by Hocoma AG in Volketswil, Switzerland. This exoskeleton enables us to both support and perturb movements of the upper extremities during motor learning, which enables us to study the neural basis of motor learning in healthy subjects and patient populations with movement disorders.

Computational Imaging

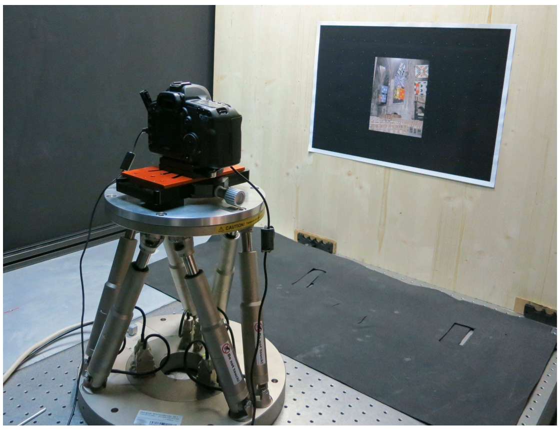


Figure 1.20: Hexapod robot with camera mounted to create repeatable camera shakes.

In computational imaging it is only of limited value to work on simulated images. Instead it is important to develop deblurring and denoising algorithms that work on real data. To be able to take such photographs for varying setups and with controlled distortions, we use several SLR and CCD cameras (including cooled and uncooled Canon 5D Mk. II, Canon 5DS R, Sony A7R Mk. II, Finger Lakes Instruments ML-16803) with various lenses. For quality assessment and quantitative comparisons we use an image quality analysis system (iQ Analyer by Image Engineering). To generate images and image sequences with controlled camera shake, we also use a Stewart platform (hexapod robot, see Figure 1.20) that allows a high accuracy for repeated movements.

In addition to usual photographic lenses we also make use of telescopes (including a Meade 12” Schmidt-Cassegrain) that allow us to work on real-world astronomical image sequences affected by turbulence. Furthermore we employ a setup with two telescopic mounts (ASA DDM60PRO, Vixen Sphinx) to study and develop reinforcement learning algorithms to improve their tracking precision.

1.4 Awards & Honors

2015

Shai Ben-David and **Ruth Urner** Best Paper Award at NIPS 2015 Workshop on "Transfer and Multitask Learning: Trends and New Perspectives".

Jakob Zscheischler: Köppen Award for his excellent doctoral dissertation (in climate and Earth system research) done at the MPI for Intelligent Systems and the MPI for Biogeochemistry in Jena.

Vinay Jayaram: Second prize in the "T3 Award" - Thesis in 3 - category PhD students. Title of the presentation: "Why can't we control computers with our minds yet?"

Leonardo Casarsa: Second prize in the "T3 Award" - Thesis in 3 - category Master students. Title of the presentation: "Can computers treat sick brains?"

Matej Balog: Winner of Code Hunt programming contest at Microsoft Research PhD Summer School 2015

Jakob Zscheischler: Otto-Hahn-Medal for his PhD Thesis "A global analysis of extreme events and consequences for the terrestrial carbon cycle", performed at the MPI for Intelligent Systems, Tübingen and the MPI for Biogeochemistry, Jena.

Katharina Muelling: Award for Extraordinary Scientific Achievement for her Ph.D. thesis by the "Freunde der TU Darmstadt".

Oliver Kroemer: Runner Up for the 2015 Georges Giralt Award (Best European Robotics Ph.D. Thesis Award)

Jan Peters : ERC Starting Grant (2015-2020) for Research on Movements of Humanoid Robotics

Philipp Hennig: Emmy-Noether-Program "Probabilistic Numerics - Probabilistic Programming for Autonomous Systems"

2014

Carl-Johann Simon-Gabriel: Google European Doctorate Fellowship 2014-2016.

M. Deisenroth (TU Darmstadt), P. Englert (Uni Stuttgart), **J. Peters**, D. Fox (U. Washington): IROS Best Cognitive Robotics Paper Award.

O. Kroemer (TU Darmstadt), H. van Hoof (TU Darmstadt), G. Neumann (TU Darmstadt), **J. Peters:** IROS Best Cognitive Robotics Paper Award Finalist

A. Abdolmaleki (Universidade de Aveiro), N. Lau (Universidade de Aveiro), G. Neumann (TU Darmstadt), **J. Peters:** 1st Place at the 3D Free Challenge of the RoboCup 2014.

Sebastian Weichwald, **Timm Meyer**, **Bernhard Schölkopf**, Tonio Ball, and **Moritz Grosse-Wentrup:** best student paper award at the 4th International Workshop on Cognitive Information Processing (CIP 2014) for: "Decoding Index Finger Position From EEG Using Random Forests".

Moritz Grosse-Wentrup and Daniel Braun: Teaching Award Winter & Summer Term 2013/14 at Graduate School of Neural Information Processing, University of Tübingen

Bernhard Schölkopf: Royal Society Milner Award 2014

2013

Manuel Gomez-Rodriguez: outstanding paper award at NIPS 2013 for the paper: "Scalable Influence Estimation in Continuous-Time Diffusion Networks"

Jan Peters: Young Investigator Award of the International Neural Networks Society (INNS) in 2013.

Jan Peters: Early Career Award of the IEEE Robotics and Automation Society in 2013.

Jens Kober: Georges Giralt Award as the Best European Robotics PhD Thesis for his PhD Thesis "Learning Motor Skills: From Algorithms to Robot Experiments".

Jonas Peters: ETH Silver Medal for his PhD thesis "Restricted Structural Equation Models for Causal Inference".

Thomas Schultz: MRM Distinguished Reviewer Award.

2012

Ulrike von Luxburg: Heisenberg Professorship

Jan Peters: Early Career Spotlight of the conference Robotics: Science & Systems (R:SS) in 2012.

Jan Peters: Frontiers of Artificial Intelligence speaker at European Conference on Artificial Intelligence (ECAI) in 2012.

Christian Daniel, Gerhard Neumann, and **Jan Peters**: IROS CoTeSys Cognitive Robotics Best Paper Award, IROS 2012 Best Student Paper Award Finalist, and IROS 2012 Best Paper Award Finalist for "Learning Concurrent Motor Skills in Versatile Solution Spaces" at the International Conference on Intelligent Robot Systems (IROS)

Stefan Harmeling: Günter Petzow Prize of the Max-Planck-Institute for Intelligent Systems

Bernhard Schölkopf: Academy Prize 2012 of the Berlin-Brandenburg Academy of Sciences and Humanities

2011

Moritz Grosse-Wentrup and **Bernhard Schölkopf:** Annual BCI Research Award 2011 for the project "What are the neurophysiological causes of performance variations in brain-computer interfacing?"

Thorsten Zander: Willumeit award for his Ph.D. thesis entitled "Utilizing Brain-Computer Interfaces for Human-Machine Systems".

Harold Christopher Burger, and **Stefan Harmeling**: DAGM 2011 Prize for "Improving Denoising Algorithms via a Multi-Scale Meta-Procedure".

Suvrit Sra: Best Paper Runner up Award in Data Mining at The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) 2011 for "Fast Projections onto L1,q-Norm Balls for Grouped Feature Selection".

Jan Peters: Dick Volz Award Best 2007 US Ph.D. Thesis Runner-Up Award.

Suvrit Sra: SIAM Outstanding Paper Prizes for his paper "The Metric Nearness Problem," which appeared in the SIAM Journal on Matrix Analysis and Applications (SIMAX).

Gustavo Camps-Valls, **Joris Mooij**, and **Bernhard Schölkopf**: were selected for the 'IEEE Geoscience and Remote Sensing Society 2011 Letters' Prize Paper Award for "Remote Sensing Feature Selection by Kernel Dependence estimation".

Bernhard Schölkopf: Annual Max Planck Research Award (shared with Sebastian Thrun, Stanford)

2010

Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause: Best Paper Award Honorable Mention at the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'2010) for the paper: "Inferring Networks of Diffusion and Influence."

Joris Mooij: Winner of two categories of the 2010 UAI Approximate Inference Challenge, 26th Conference on Uncertainty in Artificial Intelligence

Bernhard Schölkopf: Inclusion in the list of 'ISI Highly Cited Researchers'

Hirotaka Hachiya, **Jan Peters**, Mashashi Sugiyama: Young Researchers Award from IEEE Computational Intelligence Society of the Japan Chapter for their paper "Adaptive Importance Sampling with Automatic Model Selection in Reward Weighted Regression" presented at the Workshop of IEEE CIS Technical Committee on Neurocomputing (March 9-11, 2010).

Carl Edward Rasmussen and Chris Williams: 2009 DeGroot Prize for textbooks or monographs concerned with fundamental issues of statistical inference, decision theory and/or statistical applications, and chosen based on their novelty, thoroughness, timeliness, and importance of their intellectual scope: Gaussian Processes for Machine Learning, MIT Press 2006.

Oliver Kroemer, Renaud Detry, **Justus Piater**, **Jan Peters**: Best Paper Award of the 7th International Conference on Informatics in Control, Automation and Robotics (ICINCO) for: "Grasping with Vision Descriptors and Motor Primitives".

Povilas Daniusis, **Dominik Janzing**, **Joris Mooij**, **Jakob Zscheischler**, **Bastian Steudel**, **Kun Zhang**,**Bernhard Schölkof**: Best student paper award for: "Inferring Deterministic Causal Relations", 26th Conference on Uncertainty in Artificial Intelligence (UAI)

Jan Peters, Jun Morimoto, Nick Roy, Jun Nakanishi: Most Active Technical Committee Award of the IEEE Robotics and Automation Society at ICRA 2010

José M. Leiva, **Suzanne M.M. Martens**: winning team of the Machine Learning for Signal Processing (MLSP) 2010 Competition: Mind Reading

2009

Reshad Hosseini, **Matthias Bethge**: prize winning method for the highest performance in the main challenge of GREAT08 PASCAL competition (Gravitational Lensing and Accuracy Testing 2008)

Bharath Sriperumbudur, **Kenji Fukumizu**, **Arthur Gretton**, Gert Lanckriet, **Bernhard Schölkopf**: Honorable Mention, Outstanding student paper award for "Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions" at Twenty-Third Annual Conference on Neural Information Processing Systems (NIPS)

Nino Shervashidze, **Karsten Borgwardt**: Fast Subtree Kernels on Graphs, Outstanding student paper award, Twenty-Third Annual Conference on Neural Information Processing Systems (NIPS)

Christian Walder, Martin Breidt, Heinrich Bülthoff, **Bernhard Schölkopf**, Cristobal Curio: Markerless 3D Face Tracking, Runner up prize of the conference of the German Association for Pattern Recognition (DAGM)

Michael Hirsch, **Stefan Harmeling**, **Bernhard Schölkopf**: top score in 'low noise' GREAT08 PASCAL competition (Gravitational Lensing and Accuracy Testing 2008)

Stefan Harmeling, **Michael Hirsch**, **Suvrit Sra**, **Bernhard Schölkopf**: Best Poster Award for "Online Blind Deconvolution with Super-Resolution & Saturation Correction", at the 1st International Conference on Cosmology and Statistics (CosmoStats09)

Matthias Hofmann, **Florian Steinke**, Verena Scheel, **Guillaume Charpiat**, **Jason Farquhar**, Philip Aschoff, Michael Brady, **Bernhard Schölkopf** and Bernd J. Pichler: MRI-Based Attenuation Correction for PET/MRI: A Novel Approach Combining Pattern Recognition and Atlas Registration, Best Paper Awards 2008 of The Journal of Nuclear Medicine

1.5 Director profile: Bernhard Schölkopf

Bernhard Schölkopf studied Physics, Mathematics and Philosophy in Tübingen and London. In 1994 he joined Bell Labs to work on a Ph.D. with Vladimir Vapnik. Following researcher positions at GMD, Microsoft Research, and a biotech startup, Schölkopf started his lab at the Max Planck Institute for Biological Cybernetics (Tübingen) in 2002. In 2011, he became a founding director of the Max Planck Institute for Intelligent Systems.

Bernhard Schölkopf has been program chair of NIPS and COLT and is currently co-editor-in-chief of the flagship journal in machine learning (JMLR). He has been elected to the boards of the NIPS foundation and of the International Machine Learning Society. With Alex Smola, he initiated the *Machine Learning Summer Schools* series in 2002, which has meanwhile been organized, by various teams, 30 times. Many of his past students and postdocs have gone into academia (around 30 tenured or tenure-track positions) as well as to R&D labs (around 25 tenured), and he is one of the most highly cited researchers in Computer Science worldwide.[4]

Appointments

2015 – present	Co-Director of the Max Planck ETH Center for Learning Systems
2015 – present	Member of the Max Planck Campus Triumvirat, Tübingen
2011 – 2013	Managing Director, MPI for Intelligent Systems, Stuttgart & Tübingen
2011 – present	Director, Max Planck Institute for Intelligent Systems
2012 – present	Guest Professor, ETH Zürich, Computer Science Department
2010 – present	Honorarprofessor, Tübingen University, Department of Mathematics and Physics
2002 – present	Honorarprofessor, Technical University Berlin, Department of Computer Science

Awards & Honors (2009 – 2015)

2015	Overseas Visiting Scholarship, St. John's College, Cambridge, UK
2014	Royal Society Milner Award, London, UK
2013	XXVIIIth Courant Lectures, New York University
2012	Academy Prize, Berlin-Brandenburg Academy of Sciences and Humanities

[4] http://web.cs.ucla.edu/~palsberg/h-number.html

2011	Posner keynote lecturer at the NIPS Conference, Granada, Spain
2011	Max Planck Research Award, Max Planck Society
2011	Annual Brain Computer Interfaces Research Award (with Moritz Grosse-Wentrup)
2010	Inclusion in the list of ISI Highly Cited Researchers

Current Memberships

SimTECH (Excellence Cluster for Simulation Technology, Stuttgart University); Center for Integrative Neuroscience (Excellence Cluster, Tübingen University); Bernstein Focus for Neurotechnology (Freiburg/Tübingen); Bernstein Center for Computational Neuroscience (BCCN, Tübingen); German Association for Pattern Recognition (DAGM); Deutsche Mathematiker Vereinigung (DMV); Association for Computing Machinery (ACM); Institute of Electrical and Electronics Engineers (IEEE, Senior Member); European Academy of Sciences and Arts.

Committees, Service, Board Memberships

Co-founder of DALI – Data, Learning, and Inference (2015 –); Core Committee Member, MPI for Biological Cybernetics (2015 –); The Future of AI – A New York University Symposium on Science, Technology, Reason and Ethics (2016); ACM Heidelberg Laureate Forum Committee (2014–), Section Panel for Mathematics in Science and Technology, International Congress of Mathematicians (ICM) 2014; General Chair of the International Conference on Artificial Intelligence and Statistics (AISTATS) 2012; Initial Training Network for Machine Learning for Personalized Medicine (MLPM); PASCAL/PASCAL2 EU Network of Excellence; Forum Scientiarum at the University of Tübingen; Snowbird Learning workshop; Machine Learning Summer Schools (MLSS); International Machine Learning Society (IMLS); Neural Information Processing Systems Foundation (NIPS).

Review Panels

Member of the pool of experts in the process to establish the Alan Turing Institute; Centre for Doctoral Training (CDT) in Data Science, Edinburgh, UK; Italian Institute of Technology (IIT); Computer Science Department, École Normale Supérieure, Paris; Neural Computation and Adaptive Perception Program of the Canadian Institute of Advanced Research (Chair); Gatsby Computational Neuroscience Unit, Quinquennial Review 2010; Machine Learning Program, NICTA (Sydney).

Editorial Board

Journal of Machine Learning Research (JMLR, Co-editor-in-chief); International Journal of Computer Vision (2004 – 2010); Foundations and Trends in Machine Learning (since 2007); SIAM Journal on Imaging Sciences (2007 – 2012); Series *ACM Books*, Association for Computing Machinery.

Invited Talks 2009 – 2015 (Selection)

IFAC SYSID, Beijing (2015); Drawing Causal Inference from Big Data, National Academy of Sciences, Washington, D.C. (2015); International Symposium on Statistical Learning and Data Sciences, London (2015); Annual Oxford-Warwick Lectures in Statistics, Oxford (2015); Microsoft Distinguished Research Lecture, Cambridge (2015); Babbage Seminar, University of Cambridge (2015); Annual Meeting of the Max Planck Earth System Research Partnership, Weimar (2014); Symposium on Hyperspectral Imaging in Research and Engineering, Sheffield (2014); Regularization, Optimization and Kernels, Leuven (2013); Helmholtz Lecture, Humboldt-Universität Berlin (2012); International Conference on Data Mining, Brussels (2012); Maximum Entropy Conference, Munich (2012); Pattern Recognition in Neuroimaging, London (2012); Information Hiding Conference, Prague (2011); 4th Pao-Lu Hsu Lecture, Joint Statistics Lab of Microsoft Research Asia and Peking University (2010); IPMU, Dortmund (2010); SANUM, Stellenbosch (2010); Techkriti, IIT Kanpur (2010); Methods for Mining Massive Data, Copenhagen (2009).

1.6 Publications

1.6.1 Books

2014

[1] J. Kober, J. Peters. Learning Motor Skills: From Algorithms to Robot Experiments. Vol. 97. Springer Tracts in Advanced Robotics. Springer, 191 pages, 2014.

2013

[2] B. Schölkopf, Z. Luo, V. Vovk. Empirical Inference - Festschrift in Honor of Vladimir N. Vapnik. Springer, 2013.

2011

[3] S. Sra, S. Nowozin, S. Wright. Optimization for Machine Learning. Neural information processing series. MIT Press, Cambridge, MA, USA, 494 pages, 2011.

[4] H.-S. Lu, B. Schölkopf, H. Zhao. Handbook of Statistical Bioinformatics. Springer Handbooks of Computational Statistics. Springer, Berlin, Germany, 627 pages, 2011.

[5] D. Barber, A. Cemgil, S. Chiappa. Bayesian Time Series Models. Cambridge University Press, Cambridge, UK, 432 pages, 2011.

2010

[6] O. Sigaud, J. Peters. From Motor Learning to Interaction Learning in Robots. Studies in Computational Intelligence ; 264. Springer, Berlin, Germany, 538 pages, 2010.

1.6.2 Proceedings

2013

[7] M. Deisenroth, C. Szepesvári, J. Peters, editors. Proceedings of the 10th European Workshop on Reinforcement Learning (EWRL 2012). Vol. 24. JMLR Workshop and Conference Proceedings. JMLR, 173 pages, 2013.

2012

[8] E. Panagiotaki, L. O'Donnell, T. Schultz, G. Zhang, editors. Proceedings of the Workshop on Computational Diffusion MRI at MICCAI 2012 (electronic publication), 221 pages, 2012.

[9] G. Langs, I. Rish, M. Grosse-Wentrup, B. Murphy, editors. Machine Learning and Interpretation in Neuroimaging - Revised Selected and Invited Contributions, International Workshop, MLINI at NIPS 2011. Vol. LNCS 7263. Springer, Heidelberg, Germany, 266 pages, 2012.

2011

[10] S. M. Kakade, U. von Luxburg, editors. Proceedings of the 24th Annual Conference on Learning Theory (COLT 2011). Vol. 19. JMLR Workshop and Conference Proceedings. JMLR, Cambridge, MA, USA, 834 pages, 2011.

2010

[11] I. Guyon, D. Janzing, B. Schölkopf, editors. Proceedings of the Workshop: Causality: Objectives and Assessment at NIPS 2008. Vol. 6. JMLR Workshop and Conference Proceedings. JMLR, Cambridge, MA, USA, 288 pages, 2010.

1.6.3 Journal Articles

2016

[12] H. Peng, T. Huang, K. Zhang. Model Selection for Gaussian Mixture Models. *Statistica Sinica*. to appear, 2016.

[13] V. Jayaram, M. Alamgir, Y. Altun, B. Schölkopf, M. Grosse-Wentrup. Transfer Learning in Brain-Computer Interfaces. *IEEE Computational Intelligence Magazine* **11** (1): 20–31, 2016 (cited on pages 9, 30).

[14] M. Grosse-Wentrup, D. Janzing, M. Siegel, B. Schölkopf. Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach. *NeuroImage* **125**: 825–833, 2016 (cited on pages 7, 30).

[15] R. Babbar, I. Partalas, E. Gaussier, M. Amini, C. Amblard. Learning Taxonomy Adaptation in Large-scale Classification. *Journal of Machine Learning Research*. to appear, 2016.

[16] E. Klenske, M. Zeilinger, B. Schölkopf, P. Hennig. Gaussian Process Based Predictive Control for Periodic Error Correction. *IEEE Transactions on Control Systems Technology* **24** (1): 110–121, 2016 (cited on pages 7, 20).

[17] K. Zhang, Z. Wang, J. Zhang, B. Schölkopf. On estimation of functional causal models: General results and application to post-nonlinear causal model. *ACM Transactions on Intelligent Systems and Technologies* **7** (2), 2016.

2015

[18] S. Sippel, J. Zscheischler, M. Heimann, F. E. L. Otto, J. Peters, M. D. Mahecha. Quantifying changes in climate variability and extremes: Pitfalls and their overcoming. *Geophysical Research Letters* **42** (22): 9990–9998, 2015.

[19] M. R. Divine, P. Katiyar, U. Kohlhofer, L. Quintanilla-Martinez, J. A. Disselhorst, B. J. Pichler. A Population Based Gaussian Mixture Model Incorporating 18F-FDG-PET and DW-MRI Quantifies Tumor Tissue Classes. *Journal of Nuclear Medicine* **57** (3): 473–479, 2016.

[20] J. Peters, P. Bühlmann, N. Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society, Series B*. to appear, 2015 (cited on pages 6, 19).

[21] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, K. J. Resch. A quantum advantage for inferring causal structure. *Nature Physics* **11** (5): 414–420, 2015.

[22] T. Betz, R. M. Shapley, F. A. Wichmann, M. Maertens. Testing the role of luminance edges in White's illusion with contour adaptation. *Journal of Vision* **15** (11): 1–16, 2015 (cited on page 32).

[23] T. Betz, R. M. Shapley, F. A. Wichmann, M. Maertens. Noise masking of White's illusion exposes the weakness of current spatial filtering models of lightness perception. *Journal of Vision* **15** (14): 1–17, 2015 (cited on page 32).

[24] D. Janzing, B. Schölkopf. Semi-Supervised Interpolation in an Anticausal Learning Scenario. *Journal of Machine Learning Research* **16**: 1923–1948, 2015 (cited on pages 7, 19).

[25] J. F. Ramirez-Villegas, N. K. Logothetis, M. Besserve. Diversity of sharp wave-ripple LFP signatures reveals differentiated brain-wide dynamical events. *Proceedings of the National Academy of Sciences U.S.A* **112** (46): E6379–E6387, 2015 (cited on page 9).

[26] C. J. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf. Learning to Deblur. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2015 (cited on pages 23, 24).

[27] P. Melchior, E. Suchyta, E. Huff, M. Hirsch, T. Kacprzak, E. Rykoff, D. Gruen, R. Armstrong, D. Bacon, K. Bechtol et al. Mass and galaxy distributions of four massive galaxy clusters from Dark Energy Survey Science Verification data. *Monthly Notices of the Royal Astronomical Society* **449** (3): 2219–2238, 2015.

[28] I. Bezrukov, H. Schmidt, S. Gatidis, F. Mantlik, J. F. Schäfer, N. Schwenzer, B. J. Pichler. Quantitative evaluation of segmentation- and atlas- based attenuation correction for PET/MR on pediatric patients. *Journal of Nuclear Medicine* **56** (7): 1067–1074, 2015 (cited on page 25).

[29] H. F. Wehrl, I. Bezrukov, S. Wiehr, M. Lehnhoff, K. Fuchs, J. G. Mannheim, L. Quintanilla-Martinez, M. Kneilling, B. J. Pichler, A. W. Sauter. Assessment of murine brain tissue shrinkage caused by different histological fixatives using magnetic resonance and computed tomography imaging. *Histology and Histopathology* **30** (5): 601–613, 2015.

[30] C. M. Alaíz, F. Dinuzzo, S. Sra. Correlation matrix nearness and completion under observation uncertainty. *IMA Journal of Numerical Analysis* **35** (1): 325–340, 2015 (cited on page 16).

[31] R. Calandra, A. Seyfarth, J. Peters, M. Deisenroth. Bayesian Optimization for Learning Gaits under Uncertainty. *Annals of Mathematics and Artificial Intelligence*: 1–19, 2015.

[32] S. Manschitz, J. Kober, M. Gienger, J. Peters. Learning Movement Primitive Attractor Goals and Sequential Skills from Kinesthetic Demonstrations. *Robotics and Autonomous Systems* **74, Part A**: 97–107, 2015.

[33] C. Daniel, O. Kroemer, M. Viering, J. Metz, J. Peters. Active Reward Learning with a Novel Acquisition Function. *Autonomous Robots* **39** (3): 389–405, 2015.

[34] M. Besserve, S. C. Lowe, N. K. Logothetis, B. Schölkopf, S. Panzeri. Shifts of Gamma Phase across Primary Visual Cortical Sites Reflect Dynamic Stimulus-Modulated Information Transfer. *PLOS Biology* **13**: e1002257, 2015 (cited on pages 7, 9, 31).

[35] K. Muandet, B. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf. Kernel Mean Shrinkage Estimators. *Journal of Machine Learning Research*. to appear, 2016 (cited on pages 5, 14).

[36] J. Mooij, J. Peters, D. Janzing, J. Zscheischler, B. Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *Journal of Machine Learning Research*. to appear, 2016 (cited on page 18).

[37] P. Hennig, M. A. Osborne, M. Girolami. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **471** (2179), 2015 (cited on page 22).

[38] D. Foreman-Mackey, B. Montet, D. Hogg, T. Morton, D. Wang, B. Schölkopf. A systematic search for transiting planets in the K2 data. *The Astrophysical Journal* **906** (2), 2015 (cited on pages 8, 26).

[39] D. Lopez-Paz, K. Muandet, B. Recht. The Randomized Causation Coefficient. *Journal of Machine Learning*. to appear, 2016.

[40] M. Maertens, F. Wichmann, R. Shapley. Context affects lightness at the level of surfaces. *Journal of Vision* **15** (1): 1–15, 2015.

[41] R. Engbert, H. Trukenbrod, S. Barthelmé, F. Wichmann. Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision* **15** (1): 1–17, 2015 (cited on page 32).

[42] B. Schölkopf. Artificial intelligence: Learning to see and act. *Nature* **518** (7540): 486–487, 2015.

[43] B. Schölkopf, K. Muandet, K. Fukumizu, S. Harmeling, J. Peters. Computing Functions of Random Variables via Reproducing Kernel Hilbert Space Representations. *Statistics and Computing* **25** (4): 755–766, 2015 (cited on pages 5, 14, 15).

[44] C. Mariti, G. Muscolo, J. Peters, D. Puig, C. Recchiuto, C. Sighieri, A. Solanas, O. von Stryk. Developing biorobotics for veterinary research into cat movements. *Journal of Veterinary Behavior: Clinical Applications and Research* **10** (3): 248–254, 2015.

[45] A. Kupcsik, M. Deisenroth, J. Peters, L. Ai Poh, V. Vadakkepat, G. Neumann. Model-based Contextual Policy Search for Data-Efficient Generalization of Robot Skills. *Artificial Intelligence*, 2015.

[46] S. Weichwald, T. Meyer, O. Özdenizci, B. Schölkopf, T. Ball, M. Grosse-Wentrup. Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage* **110**: 48–59, 2015 (cited on pages 9, 30, 31).

[47] S. Sra. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*. published electronically, 2015.

[48] Z. Wang, A. Boularias, K. Mülling, B. Schölkopf, J. Peters. Anticipatory Action Selection for Human-Robot Table Tennis. *Artificial Intelligence*, 2015 (cited on page 8).

[49] P. Hennig. Probabilistic Interpretation of Linear Solvers. *SIAM Journal on Optimization* **25** (1): 234–260, 2015 (cited on page 21).

[50] A. Loktyushin, H. Nickisch, R. Pohmann, B. Schölkopf. Blind multirigid retrospective motion correction of MR images. *Magnetic Resonance in Medicine* **73** (4): 1457–1468, 2015 (cited on page 25).

[51] J. Zhang, K. Zhang. Likelihood and Consilience: On Forster's Counterexamples to the Likelihood Theory of Evidence. *Philosophy of Science, Supplementary Volume 2015* **82** (5): 930–940, 2015.

[52] M. Kopp, S. Harmeling, G. Schütz, B. Schölkopf, M. Fähnle. Towards denoising XMCD movies of fast magnetization dynamics using extended Kalman filter. *Ultramicroscopy* **148**: 115–122, 2015 (cited on page 7).

[53] R. Küffner, N. Zach, R. Norel, J. Hawe, D. Schoenfeld, L. Wang, G. Li, L. Fang, L. Mackey, O. Hardiman, M. Cudkowicz, A. Sherman, G. Ertaylan, M. Grosse-Wentrup, T. Hothorn, J. van Ligtenberg, J. Macke, T. Meyer, B. Schölkopf, L. Tran, et al. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nature Biotechnology* **33**: 51–57, 2015.

2014

[54] J. Zuntz, T. Kacprzak, L. Voigt, M. Hirsch, B. Rowe, S. Bridle. IM3SHAPE: Maximum likelihood galaxy shear measurement code for cosmic gravitational lensing. *Astrophysics Source Code Library* **1**: 09013, 2014.

[55] T. Kacprzak, S. Bridle, B. Rowe, L. Voigt, J. Zuntz, M. Hirsch, N. MacCrann. Sérsic galaxy models in weak lensing shape measurement: model bias, noise bias and their interaction. *Monthly Notices of the Royal Astronomical Society* **441** (3): 2528–2538, 2014.

[56] T. Schmidt-Wilcke, P. Cagnoli, P. Wang, T. Schultz, A. Lotz, W. J. Mccune, P. C. Sundgren. Diminished White Matter Integrity in Patients with Systemic Lupus Erythematosus. *NeuroImage: Clinical* **5**: 291–297, 2014 (cited on page 25).

[57] J. A. Disselhorst, I. Bezrukov, A. Kolb, C. Parl, B. J. Pichler. Principles of PET/MR Imaging. *Journal of Nuclear Medicine* **55** (6, Supplement 2): 2S–10S, 2014.

[58] J. F. Schäfer, S. Gatidis, H. Schmidt, B. Gückel, I. Bezrukov, C. A. Pfannenberg, M. Reimold, E. M., J. Fuchs, C. D. Claussen, N. F. Schwenzer. Simultaneous Whole-Body PET/MR Imaging in Comparison to PET/CT in Pediatric Oncology: Initial Results. *Radiology* **273** (1): 220–231, 2014.

[59] A. Cherian, S. Sra, V. Morellas, N. Papanikolopoulos. Efficient nearest neighbors via robust sparse hashing. *IEEE Transactions on Image Processing* **23** (8): 3646–3655, 2014.

[60] C. Persello, A. Boularias, M. Dalponte, T. Gobakken, E. Naesset, B. Schölkopf. Cost-Sensitive Active Learning With Lookahead: Optimizing Field Surveys for Remote Sensing Data Classification. *IEEE Transactions on Geoscience and Remote Sensing* **10** (52): 6652–6664, 2014.

[61] S. Martens, M. Bensch, S. Halder, J. Hill, F. Nijboer, A. Ramos-Murguialday, B. Schölkopf, N. Birbaumer, A. Gharabaghi. Epidural electrocorticography for monitoring of arousal in locked-in state. *Frontiers in Human Neuroscience* **8** (861), 2014.

[62] N. Altobelli, D. Lopez-Paz, S. Pilorz, l. Spilker, R. Morishima, S. Brooks, C. Leyrat, E. Deau, S. Edgington, A. Flandes. Two numerical models designed to reproduce Saturn ring temperatures as measured by Cassini-CIRS. *Icarus* **238**: 205–220, 2014.

[63] I. Fründ, F. Wichmann, J. Macke. Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision* **14** (7): 1–16, 2014 (cited on page 32).

[64] M. Bensch, S. Martens, S. Halder, J. Hill, F. Nijboer, A. Ramos, N. Birbaumer, M. Bodgan, B. Kotchoubey, W. Rosenstiel, B. Schölkopf, A. Gharabaghi. Assessing attention and cognitive function in completely locked-in state with event-related brain potentials and epidural electrocorticography. *Journal of Neural Engineering* **11** (2): 026006, 2014.

[65] T. Meyer, J. Peters, T. Zander, B. Schölkopf, M. Grosse-Wentrup. Predicting Motor Learning Performance from Electroencephalographic Data. *Journal of NeuroEngineering and Rehabilitation* **11:24**, 2014 (cited on page 30).

[66] C. Dann, G. Neumann, J. Peters. Policy Evaluation with Temporal Differences: A Survey and Comparison. *Journal of Machine Learning Research* **15**: 809–883, 2014.

[67] H. Ben Amor, A. Saxena, N. Hudson, J. Peters. Special issue on autonomous grasping and manipulation. *Autonomous Robots* **36** (1-2): 1–3, 2014.

[68] B. Bocsi, L. Csató, J. Peters. Indirect Robot Model Learning for Tracking Control. *Advanced Robotics* **28** (9): 589–599, 2014.

[69] M. Gomez Rodriguez, J. Leskovec, D. Balduzzi, B. Schölkopf. Uncovering the Structure and Temporal Dynamics of Information Propagation. *Network Science* **2** (1): 26–65, 2014 (cited on page 17).

[70] Z. Chen, K. Zhang, L. Chan, B. Schölkopf. Causal discovery via reproducing kernel Hilbert space embeddings. *Neural Computation* **26** (7): 1484–1517, 2014.

[71] S. Probst, S. Wiehr, F. Mantlik, H. Schmidt, A. Kolb, P. Münch, M. Delcuratolo, F. Stubenrauch, B. Pichler, T. Iftner. Evaluation of Positron Emission Tomographic Tracers for Imaging of Papillomavirus-Induced Tumors in Rabbits. *Molecular Imaging* **13** (1): 1536–0121, 2014.

[72] J. v Buttlar, J. Zscheischler, M. Mahecha. An extended approach for spatiotemporal gapfilling: dealing with large and systematic gaps in geoscientific datasets. *Nonlinear Processes in Geophysics* **21** (1): 203–215, 2014.

[73] J. Zscheischler, M. Mahecha, J. v Buttlar, S. Harmeling, M. Jung, A. Rammig, J. Randerson, B. Schölkopf, S. Seneviratne, E. Tomelleri, S. Zaehle, M. Reichstein. A few extreme events dominate global interannual variability in gross primary production. *Environmental Research Letters* **9** (3): 035001, 2014.

[74] K. Mülling, A. Boularias, B. Mohler, B. Schölkopf, J. Peters. Learning Strategies in Table Tennis using Inverse Reinforcement Learning. *Biological Cybernetics* **108** (5): 603–619, 2014.

[75] H. Schmidt, N. Schwenzer, I. Bezrukov, F. Mantlik, A. Kolb, J. Kupferschläger, B. Pichler. On the Quantification Accuracy, Homogeneity, and Stability of Simultaneous Positron Emission Tomography/Magnetic Resonance Imaging Systems. *Investigative Radiology* **49** (6): 373–381, 2014.

[76] D. Wierstra, T. Schaul, T. Glasmachers, Y. Sun, J. Peters, J. Schmidhuber. Natural Evolution Strategies. *Journal of Machine Learning Research* **15**: 949–980, 2014.

[77] J. Zscheischler, A. Michalak, M. Schwalm, M. Mahecha, D. Huntzinger, M. Reichstein, G. Berthier, P. Ciais, R. Cook, B. El-Masri, M. Huang, A. Ito, A. Jain, A. King, H. Lei, C. Lu, J. Mao, S. Peng, B. Poulter, D. Ricciuto, et al. Impact of Large-Scale Climate Extremes on Biospheric Carbon Fluxes: An Intercomparison Based on MsTMIP Data. *Global Biogeochemical Cycles*, 2014.

[78] J. Zscheischler, M. Reichstein, S. Harmeling, A. Rammig, E. Tomelleri, M. Mahecha. Extreme events in gross primary production: a characterization across continents. *Biogeosciences* **11**: 2909–2924, 2014.

[79] E. Solly, I. Schöning, S. Boch, E. Kandeler, S. Marhan, B. Michalzik, J. Müller, J. Zscheischler, S. Trumbore, M. Schrumpf. Factors controlling decomposition rates of fine root litter in temperate forests and grasslands. *Plant and Soil*, 2014.

[80] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica* **50** (3): 657–682, 2014.

[81] M. Grosse-Wentrup, B. Schölkopf. A Brain-Computer Interface Based on Self-Regulation of Gamma-Oscillations in the Superior Parietal Cortex. *Journal of Neural Engineering* **11** (5): 056015, 2014 (cited on pages 9, 30).

[82] A. Kolb, C. Parl, F. Mantlik, C. Liu, E. Lorenz, D. Renker, B. Pichler. Development of a novel depth of interaction PET detector using highly multiplexed G-APD cross-strip encoding. *Medical Physics* **41** (8), 2014.

[83] C. Wang, C. Liu, D. Roqueiro, D. Grimm, R. Schwab, C. Becker, C. Lanz, D. Weigel. Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. *Genome Research* **25** (2): 246–256, 2015.

[84] J. Peters, J. Mooij, D. Janzing, B. Schölkopf. Causal Discovery with Continuous Additive Noise Models. *Journal of Machine Learning Research* **15**: 2009–2053, 2014 (cited on pages 6, 18).

[85] P. Gerber, L. Schlaffke, S. Heba, M. Greenlee, T. Schultz, T. Schmidt-Wilcke. Juggling revisited — A voxel based morphometry study with expert jugglers. *NeuroImage* **95**: 320–325, 2014.

2013

[86] S. R. Soekadar, J. Born, N. Birbaumer, M. Bensch, S. Halder, A. R. Murguialday, A. Gharabaghi, F. Nijboer, B. Schölkopf, S. Martens. Fragmentation of Slow Wave Sleep after Onset of Complete Locked-In State. *Journal of Clinical Sleep Medicine* **9** (9): 951–953, 2013.

[87] P. Drewe, O. Stegle, L. Hartmann, A. Kahles, R. Bohnert, A. Wachter, K. M. Borgwardt, G. Rätsch. Accurate detection of differential RNA processing. *Nucleic Acids Research* **41** (10): 5189–5198, 2013.

[88] N. Fusi, C. Lippert, K. M. Borgwardt, N. D. Lawrence, O. Stegle. Detecting regulatory gene–environment interactions with unmeasured environmental factors. *Bioinformatics* **29** (11): 1382–1389, 2013.

[89] J. Zuntz, T. Kacprzak, L. Voigt, M. Hirsch, B. Rowe, S. Bridle. im3shape: a maximum likelihood galaxy shear measurement code for cosmic gravitational lensing. *Monthly Notices of the Royal Astronomical Society* **434** (2): 1604–1618, 2013.

[90] L. Bottou, J. Peters, J. Quiñonero-Candela, D. Charles, D. Chickering, E. Portugualy, D. Ray, P. Simard, E. Snelson. Counterfactual Reasoning and Learning Systems: The Example of Computational Advertising. *Journal of Machine Learning Research* **14**: 3207–3260, 2013 (cited on page 19).

[91] P. Hennig, M. Kiefel. Quasi-Newton Methods: A New Direction. *Journal of Machine Learning Research* **14** (1): 843–865, 2013 (cited on page 21).

[92] M. Maier, U. von Luxburg, M. Hein. How the result of graph clustering methods depends on the construction of the graph. *ESAIM: Probability & Statistics* **17**: 370–418, 2013.

[93] D. Balduzzi, G. Tononi. What can neurons do for their brain? Communicate selectivity with bursts. *Theory in Biosciences* **132** (1): 27–39, 2013.

[94] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, B. Schölkopf. Quantifying causal influences. *Annals of Statistics* **41** (5): 2324–2358, 2013 (cited on page 19).

[95] D. Balduzzi, P. Ortega, M. Besserve. Metabolic cost as an organizing principle for cooperative learning. *Advances in Complex Systems* **16** (02n03): 1350012, 2013 (cited on pages 10, 31).

[96] C. Persello. Interactive Domain Adaptation for the Classification of Remote Sensing Images using Active Learning. *IEEE Geoscience and Remote Sensing Letters* **10** (4): 736–740, 2013.

[97] Z. Wang, K. Mülling, M. Deisenroth, H. Ben Amor, D. Vogt, B. Schölkopf, J. Peters. Probabilistic movement modeling for intention inference in human-robot interaction. *International Journal of Robotics Research* **32** (7): 841–858, 2013.

[98] A. Loktyushin, H. Nickisch, R. Pohmann, B. Schölkopf. Blind Retrospective Motion Correction of MR Images. *Magnetic Resonance in Medicine (MRM)* **70** (6): 1608–1618, 2013 (cited on pages 8, 25).

[99] S. Barthelmé, H. Trukenbrod, R. Engbert, F. Wichmann. Modeling fixation locations using spatial point processes. *Journal of Vision* **13** (12): 1–34, 2013 (cited on page 32).

[100] A. Boularias, B. Chaib-draa. Apprenticeship Learning with Few Examples. *Neurocomputing* **104**: 83–96, 2013.

[101] S. Sra, D. Karp. The multivariate Watson distribution: Maximum-likelihood estimation and other aspects. *Journal of Multivariate Analysis* **114**: 256–269, 2013.

[102] K. Mülling, J. Kober, O. Kroemer, J. Peters. Learning to Select and Generalize Striking Movements in Robot Table Tennis. *International Journal of Robotics Research* **32** (3): 263–279, 2013.

[103] I. Bezrukov, F. Mantlik, H. Schmidt, B. Schölkopf, B. Pichler. MR-based PET Attenuation Correction for PET/MR Imaging. *Seminars in Nuclear Medicine* **43** (1): 45–59, 2013 (cited on page 8).

[104] M. Mechelke, M. Habeck. A probabilistic model for secondary structure prediction from protein chemical shifts. *Proteins: Structure, Function, and Bioinformatics* **81** (6): 984–993, 2013.

[105] M. Reichstein, M. Bahn, P. Ciais, D. Frank, M. Mahecha, S. Seneviratne, J. Zscheischler, C. Beer, N. Buchmann, D. Frank, D. Papale, A. Rammig, P. Smith, K. Thonicke, M. van der Velde, S. Vicca, A. Walz, M. Wattenbach. Climate Extremes and the Carbon Cycle. *Nature* **500**: 287–295, 2013.

[106] I. Bezrukov, H. Schmidt, F. Mantlik, N. Schwenzer, C. Brendle, B. Schölkopf, B. Pichler. MR-based Attenuation Correction Methods for Improved PET Quantification in Lesions within Bone and Susceptibility Artifact Regions. *Journal of Nuclear Medicine* **54** (10): 1768–1774, 2013 (cited on page 25).

[107] H. Gerhard, F. Wichmann, M. Bethge. How Sensitive Is the Human Visual System to the Local Statistics of Natural Images? *PLoS Computational Biology* **9** (1): e1002873, 2013.

[108] R. Goris, T. Putzeys, J. Wagemans, F. Wichmann. A neural population model for visual pattern detection. *Psychological Review* **120** (3): 472–496, 2013 (cited on pages 10, 32).

[109] F. Dinuzzo. Learning output kernels for multi-task problems. *Neurocomputing* **118**: 119–126, 2013.

[110] T. Schultz, L. Schlaffke, B. Schölkopf, T. Schmidt-Wilcke. HiFiVE: A Hilbert Space Embedding of Fiber Variability Estimates for Uncertainty Modeling and Visualization. *Computer Graphics Forum* **32** (3): 121–130, 2013 (cited on page 15).

[111] M. Bangert, P. Hennig, U. Oelfke. Analytical probabilistic modeling for radiation therapy treatment planning. *Physics in Medicine and Biology* **58** (16): 5401–5419, 2013 (cited on page 22).

[112] A. Sauter, H. Schmidt, F. Mantlik, A. Kolb, B. Federmann, C. Pfannenberg, M. Reimold, B. Pichler, W. Bethge, M. Horger. Imaging Findings and Therapy Response Monitoring in Chronic Sclerodermatous Graft-Versus-Host Disease: Preliminary Data of a Simultaneous PET/MRI Approach. *Clinical Nuclear Medicine* **38** (8): e309–e317, 2013.

[113] J. Zscheischler, M. Mahecha, S. Harmeling, M. Reichstein. Detection and attribution of large spatiotemporal extreme events in Earth observation data. *Ecological Informatics* **15**: 66–73, 2013.

[114] H. Schmidt, C. Brendle, C. Schraml, P. Martirosian, I. Bezrukov, J. Hetzel, M. Müller, A. Sauter, C. Claussen, C. Pfannenberg, N. Schwenzer. Correlation of Simultaneously Acquired Diffusion-Weighted Imaging and 2-Deoxy-[18F] fluoro-2-D-glucose Positron Emission Tomography of Pulmonary Lesions in a Dedicated Whole-Body Magnetic Resonance/Positron Emission Tomography System. *Investigative Radiology* **48** (5): 247–255, 2013.

[115] H. Wehrl, M. Hossain, K. Lankes, C. Liu, I. Bezrukov, P. Martirosian, F. Schick, G. Reischl, B. Pichler. Simultaneous PET/MR reveals Brain Function in Activated and Resting State on Metabolic, Hemodynamic and Multiple Temporal Scales. *Nature Medicine* **19**: 1184–1189, 2013.

[116] D. Varagnolo, S. Favero, F. Dinuzzo, L. Schenato, G. Pillonetto. Finding Potential Support Vectors in Separable Classification Problems. *IEEE Transactions on Neural Networks and Learning Systems* **24** (11): 1799–1813, 2013.

[117] M. Maertens, F. Wichmann. When luminance increment thresholds depend on apparent lightness. *Journal of Vision* **13** (6): 1–11, 2013 (cited on page 32).

[118] V. Schönfelder, F. Wichmann. Identification of stimulus cues in narrow-band tone-in-noise detection using sparse observer models. *Journal of the Acoustical Society of America* **134** (1): 447–463, 2013 (cited on page 32).

[119] M. Deisenroth, G. Neumann, J. Peters. A Survey on Policy Search for Robotics, Foundations and Trends in Robotics. *Foundations and Trends in Robotics* **2** (1-2): 1–142, 2013.

[120] P. Englert, A. Paraschos, J. Peters, M. Deisenroth. Probabilistic Model-based Imitation Learning. *Adaptive Behavior Journal* **21** (5): 388–403, 2013.

[121] J. Kober, D. Bagnell, J. Peters. Reinforcement Learning in Robotics: A Review. *International Journal of Robotics Research* **32** (11): 1238–1274, 2013.

[122] C.-A. Azencott, D. Grimm, M. Sugiyama, Y. Kawahara, K. Borgwardt. Efficient network-guided multi-locus association mapping with graph cuts. *Bioinformatics* **29** (13): i171–i179, 2013.

[123] D. Grimm, J. Hagmann, D. Koenig, D. Weigel, K. Borgwardt. Accurate indel prediction using paired-end short reads. *BMC Genomics* **14** (132), 2013.

[124] T. Schultz, G. Kindlmann. Open-Box Spectral Clustering: Applications to Medical Image Analysis. *IEEE Transactions on Visualization and Computer Graphics* **19** (12): 2100–2108, 2013.

[125] S. Sra. Explicit eigenvalues of certain scaled trigonometric matrices. *Linear Algebra and its Applications* **438** (1): 173–181, 2013.

2012

[126] T. Kacprzak, J. Zuntz, B. Rowe, S. Bridle, A. Refregier, A. Amara, L. Voigt, M. Hirsch. Measurement and calibration of noise bias in weak lensing galaxy shape estimation. *Monthly Notices of the Royal Astronomical Society* **427** (4): 2711–2722, 2012.

[127] T. Abbott, F. Abdalla, I. Achitouv, E. Ahn, G. Aldering, S. Allam, D. Alonso, A. Amara, J. Annis, M. Antonik et al. First SN Discoveries from the Dark Energy Survey. *The Astronomer's Telegram* **4668**: 1, 2012.

[128] T. D. Kitching, S. T. Balan, S. Bridle, N. Cantale, F. Courbin, T. Eifler, M. Gentile, M. S. S. Gill, S. Harmeling, C. Heymans, et al. Image analysis for cosmology: results from the GREAT10 Galaxy Challenge. *Monthly Notices of the Royal Astronomical Society* **423** (4): 3163–3208, 2012.

[129] C. Lampert, J. Peters. Real-time detection of colored objects in multiple camera streams with off-the-shelf hardware components. *Journal of Real-Time Image Processing* **7** (1): 31–41, 2012.

[130] H. Nickisch. glm-ie: The Generalised Linear Models Inference and Estimation Toolbox. *Journal of Machine Learning Research* **13**: 1699–1703, 2012 (cited on page 7).

[131] S. Sra. A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of Is(x). *Computational Statistics* **27** (1): 177–190, 2012.

[132] T. Zander, S. Jatzev. Context-aware brain-computer interfaces: exploring the information space of user, technical system and environment. *Journal of Neural Engineering* **9** (1): 016003, 10, 2012.

[133] M. Gomez Rodriguez, J. Leskovec, A. Krause. Inferring Networks of Diffusion and Influence. *ACM Transactions on Knowledge Discovery from Data* **5** (4:21), 2012 (cited on page 17).

[134] D. Janzing, J. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniušis, B. Steudel, B. Schölkopf. Information-geometric approach to inferring causal directions. *Artificial Intelligence* **182-183**: 1–31, 2012 (cited on pages 6, 18, 19).

[135] P. Hennig, C. Schuler. Entropy Search for Information-Efficient Global Optimization. *Journal of Machine Learning Research* **13**: 1809–1837, 2012 (cited on page 22).

[136] A. Allahverdyan, K. Hovhannisyan, D. Janzing, G. Mahler. Thermodynamic limits of dynamic cooling. *Physical Review E* **84** (4): 16, 2012.

[137] N. Hill, B. Schölkopf. An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli. *Journal of Neural Engineering* **9** (2): 026011, 2012 (cited on page 9).

[138] Y. Seldin, F. Laviolette, N. Cesa-Bianchi, J. Shawe-Taylor, P. Auer. PAC-Bayesian Inequalities for Martingales. *IEEE Transactions on Information Theory* **58** (12): 7086–7093, 2012 (cited on pages 5, 12, 29).

[139] N. Fusi, O. Stegle, N. Lawrence. Joint Modelling of Confounding Factors and Prominent Genetic Regulators Provides Increased Accuracy in Genetical Genomics Studies. *PLoS Computational Biology* **8** (1): 1–9, 2012.

[140] T. Karaletsos, O. Stegle, C. Dreyer, J. Winn, K. Borgwardt. ShapePheno: Unsupervised extraction of shape phenotypes from biological image collections. *Bioinformatics* **28** (7): 1001–1008, 2012.

[141] O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature Protocols* **7** (3): 500–507, 2012.

[142] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola. A Kernel Two-Sample Test. *Journal of Machine Learning Research* **13**: 723–773, 2012 (cited on pages 5, 14).

[143] L. Song, A. Smola, A. Gretton, J. Bedo, K. Borgwardt. Feature Selection via Dependence Maximization. *Journal of Machine Learning Research* **13**: 1393–1434, 2012 (cited on page 14).

[144] J. Kober, A. Wilhelm, E. Oztop, J. Peters. Reinforcement learning to adjust parametrized motor primitives to new situations. *Autonomous Robots* **33** (4): 361–379, 2012 (cited on pages 8, 9).

[145] M. Cavusoglu, R. Pohmann, H. Burger, K. Uludag. Regional effects of magnetization dispersion on quantitative perfusion imaging for pulsed and continuous arterial spin labeling. *Magnetic Resonance in Medicine* **69** (2): 524–530, Feb. 2013.

[146] T. Putzeys, M. Bethge, F. Wichmann, J. Wagemans, R. Goris. A New Perceptual Bias Reveals Suboptimal Population Decoding of Sensory Responses. *PLoS Computational Biology* **8** (4): 1–13, 2012.

[147] A. Cherian, S. Sra, A. Banerjee, N. Papanikolopoulos. Jensen-Bregman LogDet Divergence with Application to Efficient Similarity Search for Covariance Matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (9): 2161–2174, 2012.

[148] V. Schönfelder, F. Wichmann. Sparse regularized regression identifies behaviorally-relevant stimulus features from psychophysical data. *Journal of the Acoustical Society of America* **131** (5): 3953–3969, 2012 (cited on pages 10, 32).

[149] M. Grosse-Wentrup, B. Schölkopf. High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces. *Journal of Neural Engineering* **9** (4): 046001, 2012 (cited on pages 9, 30).

[150] D. Nguyen-Tuong, J. Peters. Online Kernel-based Learning for Task-Space Tracking Robot Control. *IEEE Transactions on Neural Networks and Learning Systems* **23** (9): 1417–1425, 2012 (cited on page 27).

[151] A. Nere, U. Olcese, D. Balduzzi, G. Tononi. A Neuromorphic Architecture for Object Recognition and Motion Anticipation Using Burst-STDP. *PLoS ONE* **7** (5): 17, 2012.

[152] A. Kolb, H. Wehrl, M. Hofmann, M. Judenhofer, L. Eriksson, R. Ladebeck, M. Lichy, L. Byars, C. Michel, H.-P. Schlemmer, M. Schmand, C. Claussen, V. Sossi, B. Pichler. Technical performance evaluation of a human brain PET/MRI system. *European Radiology* **22** (8): 1776–1788, 2012.

[153] J. Lemeire, D. Janzing. Replacing Causal Faithfulness with Algorithmic Independence of Conditionals. *Minds and Machines* **23** (2): 227–249, May 2013.

[154] B. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, G. Lanckriet. On the Empirical Estimation of Integral Probability Metrics. *Electronic Journal of Statistics* **6**: 1550–1599, 2012.

[155] T. Kam-Thong, C.-A. Azencott, L. Cayton, B. Pütz, A. Altmann, N. Karbalai, P. Sämann, B. Schölkopf, B. Müller-Myhsok, K. Borgwardt. GLIDE: GPU-Based Linear Regression for Detection of Epistasis. *Human Heredity* **73** (4): 220–236, 2012.

[156] S. Bubeck, M. Meila, U. von Luxburg. How the initialization affects the stability of the k-means algorithm. *ESAIM: Probability and Statistics* **16**: 436–452, 2012.

[157] T. Kacprzak, J. Zuntz, B. Rowe, S. Bridle, A. Refregier, A. Amara, L. Voigt, M. Hirsch. Measurement and Calibration of Noise Bias in Weak Lensing Galaxy Shape Estimation. *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 2012.

[158] S. Sra. Fast projection onto mixed-norm balls with applications. *Minining and Knowledge Discovery (DMKD)* **25** (2): 358–377, 2012.

[159] D. Kim, S. Sra, I. S. Dhillon. A non-monotonic method for large-scale non-negative least squares. *Optimization Methods and Software* **28** (5): 1012–1039, 2012.

[160] N. Logothetis, O. Eschenko, Y. Murayama, M. Augath, T. Steudel, H. Evrard, M. Besserve, A. Oeltermann. Hippocampal-Cortical Interaction during Periods of Subcortical Silence. *Nature* **491**: 547–553, 2012 (cited on pages 9, 31).

[161] B. Stranger, S. Montgomery, A. Dimas, L. Parts, O. Stegle, C. Ingle, M. Sekowska, G. Smith, D. Evans, M. Gutierrez-Arcelus et al. Patterns of cis regulatory variation in diverse human populations. *PLoS genetics* **8** (4): e1002639, 2012.

[162] O. Windram, P. Madhou, S. McHattie, C. Hill, R. Hickman, E. Cooke, D. Jenkins, C. Penfold, L. Baxter, E. Breeze, S. Kiddle, J. Rhodes, S. Atwell, D. Kliebenstein, Y-S. Kim, O. Stegle, K. Borgwardt et al. Arabidopsis defense against Botrytis cinerea: chronology and regulation deciphered by high-resolution temporal transcriptomic analysis. *The Plant Cell Online* **24** (9): 3530–3557, 2012.

[163] J. Listgarten, C. Lippert, C. Kadie, R. Davidson, E. Eskin, D. Heckerman. Improved Linear Mixed Models for Genome-Wide Association Studies. *Nature Methods* **9**: 525–526, 2012.

[164] B. Rakitsch, C. Lippert, O. Stegle, K. Borgwardt. LMM-Lasso: A Lasso Multi-Marker Mixed Model for Association Mapping with Population Structure Correction. *Bioinformatics* **29** (2): 206–214, 2012.

[165] C. Lois, I. Bezrukov, H. Schmidt, N. Schwenzer, M. Werner, J. Kupferschläger, T. Beyer. Effect of MR contrast agents on quantitative accuracy of PET in combined whole-body PET/MR imaging. *European Journal of Nuclear Medicine and Molecular Imaging* **39** (11): 1756–1766, 2012.

[166] M. Mechelke, M. Habeck. Calibration of Boltzmann distribution priors in Bayesian data analysis. *Physical Review E* **86** (6): 066705, 2012.

[167] M. Habeck. Bayesian estimation of free energies from equilibrium simulations. *Physical Review Letters* **109** (10): 5, 2012.

[168] I. Kalev, M. Mechelke, K. Kopec, T. Holder, S. Carstens, M. Habeck. CSB: A Python framework for computational structural biology. *Bioinformatics* **28** (22): 2996–2997, 2012.

[169] C. Widmer, G. Rätsch. Multitask Learning in Computational Biology. *JMLR W&CP. ICML 2011 Unsupervised and Transfer Learning Workshop* **27**: 207–216, 2012.

[170] M. Quirin, A. Loktyushin, J. Arndt, E. Küstermann, Y. Lo, J. Kuhl, L. Eggert. Existential neuroscience: a functional magnetic resonance imaging investigation of neural responses to reminders of one's mortality. *Social Cognitive and Affective Neuroscience* **7** (2): 193–198, 2012.

[171] S. Yu, H. S. Woo, H. I. Son, W. Ahn, H. Jung, D. Y. Lee, S. Y. Yi. Design of a Haptic Interface for a Gastrointestinal Endoscopy Simulation. *Advanced Robotics* **26** (18): 2115–2143, 2012.

[172] C. Persello, L. Bruzzone. Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **50** (11): 4468–4483, 2012.

2011

[173] T. Kitching, A. Amara, M. Gill, S. Harmeling, C. Heymans, R. Massey, B. Rowe, T. Schrabback, L. Voigt, S. Balan, G. Bernstein, M. Bethge, S. Bridle, F. Courbin, M. Gentile, A. Heavens, M. Hirsch, R. Hosseini, A. Kiessling, D. Kirk, et al. Gravitational Lensing Accuracy Testing 2010 (GREAT10) Challenge Handbook. *Annals of Applied Statistics* **5** (3): 2231–2263, 2011.

[174] M. Grosse-Wentrup, B. Schölkopf, J. Hill. Causal Influence of Gamma Oscillations on the Sensorimotor Rhythm. *NeuroImage* **56** (2): 837–842, 2011 (cited on pages 9, 30).

[175] M. Habeck. Statistical mechanics analysis of sparse data. *Journal of Structural Biology* **173** (3): 541–548, 2011.

[176] S. Harmeling, C. Williams. Greedy Learning of Binary Latent Trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33** (6): 1087–1097, 2011.

[177] T. Kam-Thong, D. Czamara, K. Tsuda, K. Borgwardt, C. Lewis, A. Erhardt-Lehmann, B. Hemmer, P. Rieckmann, M. Daake, F. Weber, C. Wolf, A. Ziegler, B. Pütz, F. Holsboer, B. Schölkopf, B. Müller-Myhsok. EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. *European Journal of Human Genetics* **19** (4): 465–471, 2011.

[178] J. Kober, J. Peters. Policy Search for Motor Primitives in Robotics. *Machine Learning* **84** (1-2): 171–203, 2011 (cited on pages 8, 9, 27, 28).

[179] J. Piater, S. Jodogne, R. Detry, D. Kraft, N. Krüger, O. Kroemer, J. Peters. Learning Visual Representations for Perception-Action Systems. *International Journal of Robotics Research* **30** (3): 294–307, 2011.

[180] A. Ramos Murguialday, J. Hill, M. Bensch, S. Martens, S. Halder, F. Nijboer, B. Schölkopf, N. Birbaumer, A. Gharabaghi. Transition from the locked in to the completely locked-in state: A physiological analysis. *Clinical Neurophysiology* **122** (5): 925–933, 2011.

[181] E. Georgii, K. Tsuda, B. Schölkopf. Multi-way set enumeration in weight tensors. *Machine Learning* **82** (2): 123–155, 2011.

[182] M. Grosse-Wentrup. What are the Causes of Performance Variation in Brain-Computer Interfacing? *International Journal of Bioelectromagnetism* **13** (3): 115–116, 2011.

[183] M. Grosse-Wentrup, D. Mattia, K. Oweiss. Using brain–computer interfaces to induce neural plasticity and restore function. *Journal of Neural Engineering* **8** (2): 1–5, 2011 (cited on pages 9, 30).

[184] M. Hirsch, S. Harmeling, S. Sra, B. Schölkopf. Online Multi-frame Blind Deconvolution with Super-resolution and Saturation Correction. *Astronomy & Astrophysics* **531** (A9): 11, 2011 (cited on pages 8, 23, 26).

[185] O. Kroemer, C. Lampert, J. Peters. Learning Dynamic Tactile Sensing with Robust Vision-based Training. *IEEE Transactions on Robotics* **27** (3): 545–557, 2011.

[186] D. Krusienski, M. Grosse-Wentrup, F. Galan, D. Coyle, K. Miller, E. Forney, C. Anderson. Critical issues in state-of-the-art brain–computer interface signal processing. *Journal of Neural Engineering* **8** (2): 1–8, 2011 (cited on pages 9, 30).

[187] S. Martens, J. Mooij, N. Hill, J. Farquhar, B. Schölkopf. A graphical model framework for decoding in the visual ERP-based BCI speller. *Neural Computation* **23** (1): 160–182, 2011 (cited on pages 9, 30).

[188] D. Nguyen-Tuong, J. Peters. Incremental online sparsification for model learning in real-time robot control. *Neurocomputing* **74** (11): 1859–1867, 2011.

[189] L. Parts, O. Stegle, J. Winn, R. Durbin. Joint Genetic Analysis of Gene Expression Data with Inferred Cellular Phenotypes. *PLoS Genetics* **7** (1): 1–10, 2011.

[190] M. Seeger, H. Nickisch. Large Scale Bayesian Inference and Experimental Design for Sparse Linear Models. *SIAM Journal on Imaging Sciences* **4** (1): 166–199, 2011.

[191] F. Mantlik, M. Hofmann, M. Werner, A. Sauter, J. Kupferschläger, B. Schölkopf, B. Pichler, T. Beyer. The effect of patient positioning aids on PET quantification in PET/MR imaging. *European Journal of Nuclear Medicine and Molecular Imaging* **38** (5): 920–929, 2011 (cited on page 25).

[192] M. Besserve, J. Martinerie. Extraction of functional information from ongoing brain electrical activity. *IRBM* **32** (1): 27–34, 2011.

[193] M. Chavez, M. Besserve, M. Le Van Quyen. Dynamics of excitable neural networks with heterogeneous connectivity. *Progress in Biophysics and Molecular Biology* **105** (1-2): 29–33, 2011.

[194] M. Gomez Rodriguez, J. Peters, J. Hill, B. Schölkopf, A. Gharabaghi, M. Grosse-Wentrup. Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery. *Journal of Neural Engineering* **8** (3): 1–12, 2011 (cited on pages 9, 30).

[195] M. Habeck, M. Nilges. Combining computational modeling with sparse and low-resolution data. *Journal of Structural Biology* **173** (3): 419, 2011.

[196] M. Hirsch, B. Schölkopf, M. Habeck. A Blind Deconvolution Approach for Improving the Resolution of Cryo-EM Density Maps. *Journal of Computational Biology* **18** (3): 335–346, 2011.

[197] M. Langovoy. Algebraic polynomials and moments of stochastic integrals. *Statistics & Probability Letters* **81** (6): 627–631, 2011.

[198] D. Nguyen-Tuong, J. Peters. Model learning for robot control: a survey. *Cognitive Processing* **12** (4): 319–340, 2011.

[199] R. Detry, D. Kraft, O. Kroemer, J. Peters, N. Krüger, J. Piater. Learning grasp affordance densities. *Paladyn: Journal of Behavioral Robotics* **2** (1): 1–17, 2011.

[200] J. Peters, D. Janzing, B. Schölkopf. Causal Inference on Discrete Data using Additive Noise Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33** (12): 2436–2450, 2011 (cited on page 18).

[201] M. Blaschko, J. Shelton, A. Bartels, C. Lampert, A. Gretton. Semi-supervised kernel canonical correlation analysis with application to human fMRI. *Pattern Recognition Letters* **32** (11): 1572–1583, 2011.

[202] L. Li, B. Rakitsch, K. Borgwardt. ccSVM: correcting Support Vector Machines for confounding factors in biological data classification. *Bioinformatics* **27** (13: ISMB/ECCB 2011): i342–i348, 2011.

[203] T. Kam-Thong, B. Pütz, N. Karbalai, B. Müller-Myhsok, K. Borgwardt. Epistasis detection on quantitative phenotypes by exhaustive enumeration using GPUs. *Bioinformatics* **27** (13: ISMB/ECCB 2011): i214–i221, 2011.

[204] D. Devlaminck, B. Wyns, M. Grosse-Wentrup, G. Otte, P. Santens. Multi-subject learning for common spatial patterns in motor-imagery BCI. *Computational Intelligence and Neuroscience* **2011** (217987): 1–9, 2011.

[205] M. Hofmann, I. Bezrukov, F. Mantlik, P. Aschoff, F. Steinke, T. Beyer, B. Pichler, B. Schölkopf. MRI-Based Attenuation Correction for Whole-Body PET/MRI: Quantitative Evaluation of Segmentation- and Atlas-Based Methods. *Journal of Nuclear Medicine* **52** (9): 1392–1399, 2011 (cited on page 8).

[206] C. Lippert, J. Listgarten, Y. Liu, C. Kadie, R. Davidson, D. Heckerman. FaST linear mixed models for genome-wide association studies. *Nature Methods* **8** (10): 833–835, 2011.

[207] F. Dinuzzo. Analysis of Fixed-Point and Coordinate Descent Algorithms for Regularized Kernel Methods. *IEEE Transactions on Neural Networks* **22** (10): 1576–1587, 2011.

[208] A. Ecker, P. Berens, A. Tolias, M. Bethge. The effect of noise correlations in populations of diversely tuned neurons. *Journal of Neuroscience* **31** (40): 14272–14283, 2011.

[209] K. Mülling, J. Kober, J. Peters. A biomimetic approach to robot table tennis. *Adaptive Behavior* **19** (5): 359–376, 2011.

[210] H. Hachiya, J. Peters, M. Sugiyama. Reward-Weighted Regression with Sample Reuse for Direct Policy Search in Reinforcement Learning. *Neural Computation* **23** (11): 2798–2832, 2011.

[211] D. Nguyen-Tuong, J. Peters. Model Learning in Robotics: a Survey. *Cognitive Processing*, 2011.

[212] I. Kalev, M. Habeck. HHfrag: HMM-based fragment detection using HHpred. *Bioinformatics* **27** (22): 3110–3116, 2011.

[213] C. Becker, J. Hagmann, J. Müller, D. Koenig, O. Stegle, K. Borgwardt, D. Weigel. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* **480** (7376): 245–249, 2011.

[214] J. Cao, K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, X. Wang, F. Ott, J. Müller, C. Alonso-Blanco, K. Borgwardt, K. Schmid, D. Weigel. Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nature Genetics* **43** (10): 956–963, 2011.

[215] X. Gan, O. Stegle, J. Behr, J. Steffen, P. Drewe, K. Hildebrand, R. Lyngsoe, S. Schultheiss, E. Osborne, V. Sreedharan, A. Kahles, R. Bohnert, G. Jean, P. Derwent, P. Kersey, E. Belfield, N. Harberd, E. Kemen, C. Toomajian, P. Kover, et al. Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature* **477** (7365): 419–423, 2011.

[216] B. Schölkopf. Empirical Inference. *International Journal of Materials Research* **2011** (7): 809–814, 2011.

[217] N. Shervashidze, P. Schweitzer, E. van Leeuwen, K. Mehlhorn, M. Borgwardt. Weisfeiler-Lehman Graph Kernels. *Journal of Machine Learning Research* **12**: 2539–2561, 2011.

[218] J. Peters, J. Peters, K. Mülling, Y. Altun. Reinforcement Learning with Bounded Information Loss. *AIP Conference Proceedings* **1305** (1): 365–372, 2011 (cited on page 9).

[219] I. Fründ, N. Haenel, F. Wichmann. Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision* **11** (6): 1–19, 2011.

[220] F. Dinuzzo, G. Pillonetto, G. De Nicolao. Client–Server Multitask Learning From Distributed Datasets. *IEEE Transactions on Neural Networks* **22** (2): 290–303, 2011.

[221] B. Demir, C. Persello, L. Bruzzone. Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* **49** (3): 1014–1031, 2011.

2010

[222] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, K.-R. Müller. How to Explain Individual Classification Decisions. *Journal of Machine Learning Research* **11**: 1803–1831, 2010.

[223] A. Barbero, M. Grosse-Wentrup. Biased Feedback in Brain-Computer Interfaces. *Journal of Neuro-Engineering and Rehabilitation* **7** (34): 1–4, 2010 (cited on page 9).

[224] M. Besserve, B. Schölkopf, N. Logothetis, S. Panzeri. Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *Journal of Computational Neuroscience* **29** (3): 547–566, 2010 (cited on pages 9, 31).

[225] F. Biessmann, F. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. Logothetis, K.-R. Müller. Temporal Kernel CCA and its Application in Multimodal Neuronal Data Analysis. *Machine Learning* **79** (1-2): 5–27, 2010.

[226] A. Boss, A. Kolb, M. Hofmann, S. Bisdas, T. Nägele, U. Ernemann, L. Stegger, C. Rossi, H.-P. Schlemmer, C. Pfannenberg, M. Reimold, C. Claussen, B. Pichler, U. Klose. Diffusion Tensor Imaging in a Human PET/MR Hybrid System. *Investigative Radiology* **45** (5): 270–274, 2010.

[227] A. Boss, S. Bisdas, A. Kolb, M. Hofmann, U. Ernemann, C. Claussen, C. Pfannenberg, B. Pichler, M. Reimold, L. Stegger. Hybrid PET/MRI of Intracranial Masses: Initial Experiences and Comparison to PET/CT. *Journal of Nuclear Medicine* **51** (8): 1198–1205, 2010.

[228] S. Bridle, S. Balan, M. Bethge, M. Gentile, S. Harmeling, C. Heymans, M. Hirsch, R. Hosseini, M. Jarvis, D. Kirk, T. Kitching, K. Kuijken, A. Lewis, S. Paulin-Henriksson, B. Schölkopf, M. Velander, L. Voigt, D. Witherick, A. Amara, G. Bernstein, et al. Results of the GREAT08 Challenge: An image analysis competition for cosmological lensing. *Monthly Notices of the Royal Astronomical Society* **405** (3): 2044–2061, 2010.

[229] G. Camps-Valls, J. Mooij, B. Schölkopf. Remote Sensing Feature Selection by Kernel Dependence Estimation. *IEEE Geoscience and Remote Sensing Letters* **7** (3): 587–591, 2010.

[230] G. Camps-Valls, N. Shervashidze, K. Borgwardt. Spatio-Spectral Remote Sensing Image Classification With Graph Kernels. *IEEE Geoscience and Remote Sensing Letters* **7** (4): 741–745, 2010.

[231] D. Corfield. Varieties of Justification in Machine Learning. *Minds and Machines* **20** (2): 291–301, 2010.

[232] D. Görür, C. Rasmussen. Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution. *Journal of Computer Science and Technology* **25** (4): 653–664, 2010 (cited on page 7).

[233] D. Janzing. On the Entropy Production of Time Series with Unidirectional Linearity. *Journal of Statistical Physics* **138** (4-5): 767–779, 2010.

[234] D. Janzing, B. Schölkopf. Causal Inference Using the Algorithmic Markov Condition. *IEEE Transactions on Information Theory* **56** (10): 5168–5194, 2010 (cited on pages 6, 18).

[235] D. Janzing, B. Steudel. Justifying Additive Noise Model-Based Causal Discovery via Algorithmic Information Theory. *Open Systems and Information Dynamics* **17** (2): 189–212, 2010.

[236] D. Kim, S. Sra, I. Dhillon. Tackling Box-Constrained Optimization via a New Projected Quasi-Newton Approach. *SIAM Journal on Scientific Computing* **32** (6): 3548–3563, 2010.

[237] K. Kim, Y. Kwon. Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32** (6): 1127–1133, 2010.

[238] J. Kober, J. Peters. Imitation and Reinforcement Learning. *IEEE Robotics and Automation Magazine* **17** (2): 55–62, 2010 (cited on pages 8, 27).

[239] O. Kroemer, R. Detry, J. Piater, J. Peters. Combining active learning and reactive control for robot grasping. *Robotics and Autonomous Systems* **58** (9): 1105–1116, 2010 (cited on page 8).

[240] M. Lázaro-Gredilla, J. Quiñonero-Candela, C. Rasmussen, A. Figueiras-Vidal. Sparse Spectrum Gaussian Process Regression. *Journal of Machine Learning Research* **11**: 1865–1881, 2010 (cited on page 7).

[241] C. Lippert, Z. Ghahramani, K. Borgwardt. Gene function prediction from synthetic lethality networks via ranking on demand. *Bioinformatics* **26** (7): 912–918, 2010.

[242] J. Macke, F. Wichmann. Estimating predictive stimulus features from psychophysical data: The decision image technique applied to human faces. *Journal of Vision* **10** (5:22): 1–24, 2010 (cited on page 10).

[243] S. Martens, J. Leiva. A generative model approach for decoding in the visual event-related potential-based brain-computer interface speller. *Journal of Neural Engineering* **7** (2): 1–10, 2010.

[244] M. Mechelke, M. Habeck. Robust probabilistic superposition and comparison of protein structures. *BMC Bioinformatics* **11** (363): 1–13, 2010.

[245] J. Mooij. libDAI: A Free and Open Source C++ Library for Discrete Approximate Inference in Graphical Models. *Journal of Machine Learning Research* **11**: 2169–2173, 2010.

[246] T. Morimura, E. Uchibe, J. Yoshimoto, J. Peters, K. Doya. Derivatives of Logarithmic Stationary Distributions for Policy Gradient Reinforcement Learning. *Neural Computation* **22** (2): 342–376, 2010.

[247] J. Peters. Policy gradient methods. *Scholarpedia* **5** (11): 3698, 2010.

[248] J. Peters, J. Kober, S. Schaal. Algorithmen zum Automatischen Erlernen von Motorfähigkeiten. *at - Automatisierungstechnik* **58** (12): 688–694, 2010.

[249] C. Rasmussen, H. Nickisch. Gaussian Processes for Machine Learning (GPML) Toolbox. *Journal of Machine Learning Research* **11**: 3011–3015, 2010 (cited on page 7).

[250] M. Seeger, H. Nickisch, R. Pohmann, B. Schölkopf. Optimization of k-Space Trajectories for Compressed Sensing by Bayesian Experimental Design. *Magnetic Resonance in Medicine* **63** (1): 116–126, 2010.

[251] F. Sehnke, C. Osendorfer, T. Rückstiess, A. Graves, J. Peters, J. Schmidhuber. Parameter-exploring policy gradients. *Neural Networks* **21** (4): 551–559, 2010.

[252] Y. Seldin, N. Tishby. PAC-Bayesian Analysis of Co-clustering and Beyond. *Journal of Machine Learning Research* **11**: 3595–3646, 2010.

[253] B. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, G. Lanckriet. Hilbert Space Embeddings and Metrics on Probability Measures. *Journal of Machine Learning Research* **11**: 1517–1561, 2010 (cited on pages 5, 15).

[254] O. Stegle, K. Denby, E. Cooke, D. Wild, Z. Ghahramani, K. Borgwardt. A Robust Bayesian Two-Sample Test for Detecting Intervals of Differential Gene Expression in Microarray Time Series. *Journal of Computational Biology* **17** (3): 355–367, 2010.

[255] O. Stegle, L. Parts, R. Durbin, J. Winn. A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression Levels Greatly Increases Power in eQTL Studies. *PLoS Computational Biology* **6** (5): 1–11, 2010.

[256] O. Stegle, P. Drewe, R. Bohnert, K. Borgwardt, G. Rätsch. Statistical Tests for Detecting Differential RNA-Transcript Expression from Read Counts. *Nature Precedings* **2010**: 1–11, 2010.

[257] F. Steinke, M. Hein, B. Schölkopf. Nonparametric Regression between General Riemannian Manifolds. *SIAM Journal on Imaging Sciences* **3** (3): 527–563, 2010.

[258] T. Tuytelaars, C. Lampert, M. Blaschko, W. Buntine. Unsupervised Object Discovery: A Comparison. *International Journal of Computer Vision* **88** (2): 284–302, 2010.

[259] U. von Luxburg. Clustering stability: an overview. *Foundations and Trends in Machine Learning* **2** (3): 235–274, 2010.

[260] D. Wierstra, A. Förster, J. Peters, J. Schmidhuber. Recurrent Policy Gradients. *Logic Journal of the IGPL* **18** (5): 620–634, 2010.

[261] H. Wong, O. Stegle, S. Rodgers, J. Huppert. A toolbox for predicting G-quadruplex formation and stability. *Journal of Nucleic Acids* **2010** (564946): 1–6, 2010.

[262] K. Zhang, L.-W. Chan. Convolutive blind source separation by efficient blind deconvolution and minimal filter distortion. *Neurocomputing* **73** (13-15): 2580–2588, 2010.

[263] J.-P. Armache, A. Jarasch, A. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. Wilson, R. Beckmann. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **107** (46): 19748–19753, 2010.

[264] J.-P. Armache, A. Jarasch, A. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. Wilson, R. Beckmann. Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **107** (46): 19754–19759, 2010.

[265] S. Vishwanathan, N. Schraudolph, R. Kondor, K. Borgwardt. Graph Kernels. *Journal of Machine Learning Research* **11**: 1201–1242, 2010.

[266] M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A. Smola, L. Song, P. Yu, X. Yan, K. Borgwardt. Discriminative frequent subgraph mining with optimality guarantees. *Journal of Statistical Analysis and Data Mining* **3** (5): 302–318, 2010.

[267] F. Wichmann, J. Drewes, P. Rosas, K. Gegenfurtner. Animal detection in natural scenes: Critical features revisited. *Journal of Vision* **10** (4): 1–27, 2010.

[268] A. Gretton, L. Györfi. Consistent Nonparametric Tests of Independence. *Journal of Machine Learning Research* **11**: 1391–1423, 2010 (cited on page 14).

[269] C. Widmer, N. Toussaint, Y. Altun, G. Rätsch. Inferring latent task structure for Multitask Learning by Multiple Kernel Learning. *BMC Bioinformatics* **11 Suppl 8**: S5, 2010.

[270] A. Hyvärinen, K. Zhang, S. Shimizu, P. Hoyer. Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity. *Journal of Machine Learning Research* **11**: 1709–1731, 2010.

[271] D. Lee, J. Lee. Dynamic Dissimilarity Measure for Support-Based Clustering. *IEEE Transactions on Knowledge and Data Engineering* **22** (6): 900–905, 2010.

[272] G. Pillonetto, F. Dinuzzo, G. De Nicolao. Bayesian Online Multitask Learning of Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32** (2): 193–205, 2010.

[273] C. Persello, L. Bruzzone. A Novel Protocol for Accuracy Assessment in Classification of Very High Resolution Images. *IEEE Transactions on Geoscience and Remote Sensing* **48** (3): 1232–1244, 2010.

2009

[274] H. Shin, K. Tsuda, B. Schölkopf. Protein Functional Class Prediction With a Combined Graph. *Expert Systems with Applications* **36** (2): 3284–3292, 2009.

[275] A. Allahverdyan, D. Janzing, G. Mahler. Thermodynamic efficiency of information and heat flow. *Journal of Statistical Mechanics: Theory and Experiment* **2009** (09): P09011, 2009.

[276] B. Bardiaux, A. Bernard, W. Rieping, M. Habeck, T. Malliavin, M. Nilges. Influence of Different Assignment Conditions on the Determination of Symmetric Homo-dimeric Structures with ARIA. *Proteins* **75** (3): 569–585, 2009.

[277] S. Bubeck, U. von Luxburg. Nearest Neighbor Clustering: A Baseline Method for Consistent Clustering with Arbitrary Objective Functions. *Journal of Machine Learning Research* **10**: 657–698, 2009.

[278] D. Corfield, B. Schölkopf, V. Vapnik. Falsificationism and Statistical Learning Theory: Comparing the Popper and Vapnik-Chervonenkis Dimensions. *Journal for General Philosophy of Science* **40** (1): 51–58, 2009.

[279] M. Deisenroth, C. Rasmussen, J. Peters. Gaussian Process Dynamic Programming. *Neurocomputing* **72** (7-9): 1508–1524, 2009 (cited on page 29).

[280] S. Dietmann, E. Georgii, A. Antonov, K. Tsuda, H.-W. Mewes. The DICS repository: module-assisted analysis of disease-related gene lists. *Bioinformatics* **25** (6): 830–831, 2009.

[281] E. Georgii, S. Dietmann, T. Uno, P. Pagel, K. Tsuda. Enumeration of condition-dependent dense modules in protein interaction networks. *Bioinformatics* **25** (7): 933–940, 2009.

[282] A. Graf, O. Bousquet, G. Rätsch, B. Schölkopf. Prototype Classification: Insights from Machine Learning. *Neural Computation* **21** (1): 272–300, 2009.

[283] M. Grosse-Wentrup, C. Liefhold, K. Gramann, M. Buss. Beamforming in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Biomedical Engineering* **56** (4): 1209–1219, 2009 (cited on page 9).

[284] M. Habeck. Generation of three-dimensional random rotations in fitting and matching problems. *Computational Statistics* **24** (4): 719–731, 2009.

[285] H. Hachiya, T. Akiyama, M. Sugiyama, J. Peters. Adaptive Importance Sampling for Value Function Approximation in Off-policy Reinforcement Learning. *Neural Networks* **22** (10): 1399–1410, 2009.

[286] S. Harmeling. Inferring textual entailment with a probabilistically sound calculus. *Natural Language Engineering* **15** (4): 459–477, 2009.

[287] P. Haselager, R. Vlek, J. Hill, F. Nijboer. A note on ethical aspects of BCI. *Neural Networks* **22** (9): 1352–1357, 2009.

[288] M. Hofmann, B. Pichler, B. Schölkopf, T. Beyer. Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques. *European Journal of Nuclear Medicine and Molecular Imaging* **36** (Supplement 1): 93–104, 2009.

[289] F. Jäkel, B. Schölkopf, F. Wichmann. Does Cognitive Science Need Kernels? *Trends in Cognitive Sciences* **13** (9): 381–388, 2009 (cited on pages 10, 32).

[290] W. Kienzle, M. Franz, B. Schölkopf, F. Wichmann. Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision* **9** (5:7): 1–15, 2009 (cited on page 10).

[291] S. Klöppel, C. Chu, G. Tan, B. Draganski, H. Johnson, J. Paulsen, W. Kienzle, S. Tabrizi, J. Ashburner, R. Frackowiak. Automatic detection of preclinical neurodegeneration: Presymptomatic Huntington disease. *Neurology* **72** (5): 426–431, 2009.

[292] C. Lampert. Kernel Methods in Computer Vision. *Foundations and Trends in Computer Graphics and Vision* **4** (3): 193–285, 2009.

[293] C. Lampert, M. Blaschko. Structured prediction by joint kernel support estimation. *Machine Learning* **77** (2-3): 249–269, 2009.

[294] C. Lampert, M. Blaschko, T. Hofmann. Efficient Subwindow Search: A Branch and Bound Framework for Object Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31** (12): 2129–2142, 2009.

[295] D. Lee, K.-H. Jung, J. Lee. Constructing Sparse Kernel Machines Using Attractors. *IEEE Transactions on Neural Networks* **20** (4): 721–729, 2009.

[296] J. Macke, P. Berens, A. Ecker, A. Tolias, M. Bethge. Generating Spike Trains with Specified Correlation Coefficients. *Neural Computation* **21** (2): 397–423, 2009.

[297] M. Maier, M. Hein, U. von Luxburg. Optimal construction of k-nearest-neighbor graphs for identifying noisy clusters. *Theoretical Computer Science* **410** (19): 1749–1764, 2009.

[298] S. Martens, N. Hill, J. Farquhar, B. Schölkopf. Overlap and refractory effects in a Brain-Computer Interface speller based on the visual P300 Event-Related Potential. *Journal of Neural Engineering* **6** (2): 1–9, 2009 (cited on page 9).

[299] D. Nguyen-Tuong, M. Seeger, J. Peters. Model Learning with Local Gaussian Process Regression. *Advanced Robotics* **23** (15): 2015–2034, 2009.

[300] J. Peters, A. Ng. Guest editorial: Special issue on robot learning, Part A. *Autonomous Robots* **27** (1): 1–2, 2009.

[301] J. Peters, A. Ng. Guest editorial: Special issue on robot learning, Part B. *Autonomous Robots* **27** (2): 91–92, 2009.

[302] J. Peters, J. Kober. Policy Search for Motor Primitives. *KI - Zeitschrift Künstliche Intelligenz* **23** (3): 38–40, 2009.

[303] J. Peters, J. Morimoto, R. Tedrake, N. Roy. Robot Learning. *IEEE Robotics and Automation Magazine* **16** (3): 19–20, 2009.

[304] C. Rasmussen, B. de la Cruz, Z. Ghahramani, D. Wild. Modeling and Visualizing Uncertainty in Gene Expression Clusters using Dirichlet Process Mixtures. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **6** (4): 615–628, 2009.

[305] H. Shen, S. Jegelka, A. Gretton. Fast Kernel-Based Independent Component Analysis. *IEEE Transactions on Signal Processing* **57** (9): 3498–3511, 2009.

[306] U. von Luxburg, V. Franz. A Geometric Approach to Confidence Sets for Ratios: Fieller's Theorem, Generalizations, and Bootstrap. *Statistica Sinica* **19** (3): 1095–1117, 2009.

[307] C. Parker, Y. Altun, P. Tadepalli. Guest editorial: special issue on structured prediction. *Machine Learning* **77** (2-3): 161–164, 2009.

[308] R. Goris, F. Wichmann, G. Henning. A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision* **9** (7): 1–22, 2009.

[309] S. Djuranovic, M. Hartmann, M. Habeck, A. Ursinus, P. Zwickl, J. Martin, A. Lupas, K. Zeth. Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. *Molecular Cell* **34** (5): 580–590, 2009.

[310] F. Wichmann, W. Kienzle, B. Schölkopf, M. Franz. Non-linear System Identification: Visual Saliency Inferred from Eye-Movement Data. *Journal of Vision* **9** (8): article 32, 2009.

[311] A. Gretton, K. Fukumizu, B. Sriperumbudur. Discussion of: Brownian Distance Covariance. *The Annals of Applied Statistics* **3** (4): 1285–1294, 2009.

[312] G. Schweikert, A. Zien, G. Zeller, J. Behr, C. Dieterich, C. Ong, P. Philips, F. De Bona, L. Hartmann, A. Bohlen, N. Krüger, S. Sonnenburg, G. Rätsch. mGene: accurate SVM-based gene finding with an application to nematode genomes. *Genome Research* **19** (11): 2133–43, 2009 (cited on page 10).

[313] G. Schweikert, J. Behr, A. Zien, G. Zeller, C. Ong, S. Sonnenburg, G. Rätsch. mGene.web: a web service for accurate computational gene finding. *Nucleic Acids Research* **37**: W312–6, 2009 (cited on page 10).

1.6.4 Conference Papers

2016

[314] D. Lopez-Paz, B. Schölkopf, L. Bottou, V. Vapnik. Unifying distillation and privileged information. In *International Conference on Learning Representations (ICLR 2016)*, 2016.

[315] D. Büchler, H. Ott, J. Peters. A Lightweight Robotic Arm with Pneumatic Muscles for Robot Learning. In *IEEE International Conference on Robotics and Automation (ICRA 2016)*, 2016.

[316] R. Babbar, K. Muandet, B. Schölkopf. A Scalable Mixed-norm Approach for Learning Lightweight Models in Large-scale Classification. In *Proceedings of the 2016 SIAM International Conference on Data Mining (SDM 2016)*, 2016 (cited on page 16).

2015

[317] D. Rothenhäusler, C. Heinze, J. Peters, N. Meinshausen. BACKSHIFT: Learning causal cyclic graphs from unknown shift interventions. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 1513–1521, 2015.

[318] A. Loktyushin, C. Schuler, K. Scheffler, B. Schölkopf. Retrospective motion correction of magnitude-input MR images. In *Proceedings of the Workshop on Machine Learning meets Medical Imaging at ICML 2015*. Vol. 9487. Lecture Notes in Computer Science, pages 3–12, 2015.

[319] A. Ibarra Chaoul, M. Grosse-Wentrup. Is Breathing Rate a Confounding Variable in Brain-Computer Interfaces (BCIs) Based on EEG Spectral Power? In *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2015)*, pages 1079–1082, 2015.

[320] M. Hirsch, B. Schölkopf. Self-calibration of optical lenses. In *IEEE International Conference on Computer Vision (ICCV 2015)*, pages 612–620, 2015 (cited on pages 8, 23, 24).

[321] M. Mahsereci, P. Hennig. Probabilistic Line Searches for Stochastic Optimization. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 181–189, 2015 (cited on page 22).

[322] A. Wahrburg, S. Zeiss, B. Matthias, J. Peters, H. Ding. Combined Pose-Wrench and State Machine Representation for Modeling Robotic Assembly Skills. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS 2015)*, pages 852–857, 2015.

[323] F. Veiga, H. van Hoof, J. Peters, T. Hermans. Stabilizing Novel Objects by Learning to Predict Tactile Slip. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS 2015)*, pages 5065–5072, 2015.

[324] E. Rueckert, J. Mundo, A. Paraschos, J. Peters, G. Neumann. Extracting Low-Dimensional Control Variables for Movement Primitives. In *IEEE International Conference on Robotics and Automation (ICRA 2015)*, pages 1511–1518, 2015.

[325] S. Parisi, H. Abdulsamad, A. Paraschos, C. Daniel, J. Peters. Reinforcement Learning vs Human Programming in Tetherball Robot Games. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS 2015)*, pages 6428–6434, 2015.

[326] A. Paraschos, E. Rueckert, J. Peters, G. Neumann. Model-Free Probabilistic Movement Primitives for Physical Interaction. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS 2015)*, pages 2860–2866, 2015.

[327] S. Manschitz, J. Kober, M. Gienger, J. Peters. Probabilistic Progress Prediction and Sequencing of Concurrent Movement Primitives. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS 2015)*, pages 449–455, 2015.

[328] R. Lioutikov, G. Neumann, G. Maeda, J. Peters. Probabilistic Segmentation Applied to an Assembly Task. In *15th IEEE-RAS International Conference on Humanoid Robots*, pages 533–540, 2015.

[329] S. Leischnig, S. Luettgen, O. Kroemer, J. Peters. A Comparison of Contact Distribution Representations for Learning to Predict Object Interactions. In *15th IEEE-RAS International Conference on Humanoid Robots*, pages 616–622, 2015.

[330] O. Kroemer, C. Daniel, G. Neumann, H. van Hoof, J. Peters. Towards Learning Hierarchical Skills for Multi-Phase Manipulation Tasks. In *IEEE International Conference on Robotics and Automation (ICRA 2015)*, pages 1503–1510, 2015.

[331] O. Koc, G. Maeda, G. Neumann, J. Peters. Optimizing Robot Striking Movement Primitives with Iterative Learning Control. In *15th IEEE-RAS International Conference on Humanoid Robots*, pages 80–87, 2015.

[332] J. Hoelscher, J. Peters, T. Hermans. Evaluation of Interactive Object Recognition with Tactile Sensing. In *15th IEEE-RAS International Conference on Humanoid Robots*, pages 310–317, 2015.

[333] L. Fritsche, F. Unverzagt, J. Peters, R. Calandra. First-Person Tele-Operation of a Humanoid Robot. In *15th IEEE-RAS International Conference on Humanoid Robots*, pages 997–1002, 2015.

[334] M. Ewerton, G. Neumann, R. Lioutikov, H. Ben Amor, J. Peters, G. Maeda. Learning multiple collaborative tasks with a mixture of Interaction Primitives. In *IEEE International Conference on Robotics and Automation (ICRA 2015)*, pages 1535–1542, 2015.

[335] M. Ewerton, G. Maeda, J. Peters, G. Neumann. Learning Motor Skills from Partially Observed Movements Executed at Different Speeds. In *IEEE/RSJ Conference on Intelligent Robots and Systems (IROS 2015)*, pages 456–463, 2015.

[336] R. Calandra, S. Ivaldi, M. Deisenroth, E. Rückert, J. Peters. Learning Inverse Dynamics Models with Contacts. In *IEEE International Conference on Robotics and Automation (ICRA 2015)*, pages 3186–3191, 2015.

[337] R. Calandra, S. Ivaldi, M. Deisenroth, J. Peters. Learning Torque Control in Presence of Contacts using Tactile Sensing from Robot Skin. In *15th IEEE-RAS International Conference on Humanoid Robots*, pages 690–695, 2015.

[338] A. Abdolmaleki, J. Peters, G. Neumann. Model-Based Relative Entropy Stochastic Search. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 3523–3531, 2015.

[339] P. Awasthi, M. Balcan, N. Haghtalab, R. Urner. Efficient Learning of Linear Separators under Bounded Noise. In *Proceedings of the 28th Conference on Learning Theory (COLT 2015)*. Vol. 40. JMLR Workshop and Conference Proceedings, pages 167–190, 2015 (cited on page 12).

[340] S. Kpotufe, R. Urner, S. Ben-David. Hierarchical Label Queries with Data-Dependent Partitions. In *Proceedings of the 28th Conference on Learning Theory (COLT 2015)*. Vol. 40. JMLR Workshop and Conference Proceedings, pages 1176–1189, 2015 (cited on pages 4, 12).

[341] C. Berlind, R. Urner. Active Nearest Neighbors in Changing Environments. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*. Vol. 37. JMLR Workshop and Conference Proceedings, pages 1870–1879, 2015 (cited on pages 4, 12).

[342] I. Tolstikhin, N. Zhivotovskiy, G. Blanchard. Permutational Rademacher Complexity: a New Complexity Measure for Transductive Learning. In *Proceedings of the 26th International Conference on Algorithmic Learning Theory (ALT 2015)*. Vol. 9355. Lecture Notes in Computer Science, pages 209–223, 2015 (cited on pages 4, 12).

[343] T. Fomina, M. R. Hohmann, B. Schölkopf, M. Grosse-Wentrup. Identification of the Default Mode Network with Electroencephalography. In *Proceedings of the 37th IEEE Conference for Engineering in Medicine and Biology (EMBC 2015)*, pages 7566–7569, 2015.

[344] A. Marco, P. Hennig, J. Bohg, S. Schaal, S. Trimpe. Automatic LQR Tuning Based on Gaussian Process Optimization: Early Experimental Results. In *Machine Learning in Planning and Control of Robot Motion Workshop at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015)*, 2015.

[345] M. Khatami, T. Schmidt-Wilcke, P. Sundgren, A. Abbasloo, B. Schölkopf, T. Schultz. BundleMAP: Anatomically Localized Features from dMRI for Detection of Disease. In *6th International Workshop on Machine Learning in Medical Imaging (MLMI 2015)*. Vol. 9352. Lecture Notes in Computer Science, pages 52–60, 2015 (cited on page 25).

[346] M. R. Hohmann, T. Fomina, V. Jayaram, N. Widmann, C. Förster, J. Müller vom Hagen, M. Synofzik, B. Schölkopf, L. Schöls, M. Grosse-Wentrup. A Cognitive Brain-Computer Interface for Patients with Amyotrophic Lateral Sclerosis. In *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2015)*, pages 3187–3191, 2015 (cited on pages 9, 30).

[347] Y. Huang, B. Schölkopf, J. Peters. Learning Optimal Striking Points for A Ping-Pong Playing Robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015)*, pages 4587–4592, 2015 (cited on page 27).

[348] V. Jayaram, N. Widmann, C. Förster, T. Fomina, M. R. Hohmann, J. Müller vom Hagen, M. Synofzik, B. Schölkopf, L. Schöls, M. Grosse-Wentrup. Brain-Computer Interfacing in Amyotrophic Lateral Sclerosis: Implications of a Resting-State EEG Analysis. In *Proceedings of the 37th IEEE Conference for Engineering in Medicine and Biology (EMBC 2015)*, pages 6979–6982, 2015.

[349] B. Schölkopf, D. Hogg, D. Wang, D. Foreman-Mackey, D. Janzing, C.-J. Simon-Gabriel, J. Peters. Removing systematic errors for exoplanet search via latent causes. In *Proceedings of The 32nd International Conference on Machine Learning (ICML 2015)*. Vol. 37. JMLR Workshop and Conference Proceedings, pages 2218–2226, 2015 (cited on pages 7, 8, 19, 26).

[350] P. Spirtes, K. Zhang. Recent Methodological Advances in Causal Discovery and Inference. In *15th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2015)*, pages 23–35, 2015.

[351] P. Geiger, K. Zhang, B. Schölkopf, M. Gong, D. Janzing. Causal Inference by Identification of Vector Autoregressive Processes with Hidden Components. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*. Vol. 37. JMLR Workshop and Conference Proceedings, pages 1917–1925, 2015 (cited on page 19).

[352] M. Gong, K. Zhang, B. Schölkopf, D. Tao, P. Geiger. Discovering Temporal Causal Relations from Subsampled Data. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*. Vol. 37. JMLR Workshop and Conference Proceedings, pages 1898–1906, 2015.

[353] N. Shajarisales, D. Janzing, B. Schölkopf, M. Besserve. Telling cause from effect in deterministic linear dynamical systems. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*. Vol. 37. JMLR Workshop and Conference Proceedings, pages 285–294, 2015 (cited on pages 9, 18, 31).

[354] S. Hauberg, M. Schober, M. Liptrot, P. Hennig, A. Feragen. A Random Riemannian Metric for Probabilistic Shortest-Path Tractography. In *18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2015)*. Vol. 9349. Lecture Notes in Computer Science, pages 597–604, 2015 (cited on page 22).

[355] K. Zhang, J. Zhang, B. Schölkopf. Distinguishing Cause from Effect Based on Exogeneity. In *15th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2015)*, pages 261–271, 2015.

[356] D. Lopez-Paz, K. Muandet, B. Schölkopf, I. Tolstikhin. Towards a Learning Theory of Cause-Effect Inference. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*. Vol. 37. JMLR Workshop and Conference Proceedings, pages 1452–1461, 2015 (cited on pages 5, 7, 12, 14).

[357] B. Huang, K. Zhang, B. Schölkopf. Identification of Time-Dependent Causal Model: A Gaussian Process Treatment. In *24th International Joint Conference on Artificial Intelligence, Machine Learning Track (IJCAI15)*, pages 3561–3568, 2015.

[358] E. Sgouritsa, D. Janzing, P. Hennig, B. Schölkopf. Inference of Cause and Effect with Unsupervised Inverse Regression. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS 2015)*. Vol. 38. JMLR Workshop and Conference Proceedings, pages 847–855, 2015 (cited on page 18).

[359] H. van Hoof, J. Peters, G. Neumann. Learning of Non-Parametric Control Policies with High-Dimensional State Features. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS 2015)*. Vol. 38. JMLR Workshop and Conference Proceedings, pages 995–1003, 2015.

[360] K. Zhang, M. Gong, B. Schölkopf. Multi-Source Domain Adaptation: A Causal View. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*, pages 3150–3157, 2015 (cited on page 6).

2014

[361] D. Lang, D. Hogg, B. Schölkopf. Towards building a Crowd-Sourced Sky Map. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS 2014)*. Vol. 33. JMLR Workshop and Conference Proceedings, pages 549–557, 2014 (cited on pages 8, 26).

[362] A. W. Yu, W. Ma, Y. Yu, J. Carbonell, S. Sra. Efficient Structured Matrix Rank Minimization. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pages 1350–1358, 2014.

[363] L. Pickup, P. Zheng, W. Donglai, S. YiChang, Z. Changshui, A. Zisserman, B. Schölkopf, W. Freeman. Seeing the Arrow of Time. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2043–2050, 2014 (cited on page 32).

[364] M. Gomez Rodriguez, L. Song, B. Schölkopf. Open Problem: Finding Good Cascade Sampling Processes for the Network Inference Problem. In *Proceedings of the 27th Conference on Learning Theory (COLT 2014)*. Vol. 35. JMLR Workshop and Conference Proceedings, pages 1276–1279, 2014.

[365] M. Wytock, S. Sra, J. Z. Kolter. Fast Newton methods for the group fused lasso. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, pages 888–897, 2014 (cited on page 16).

[366] S. Azadi, S. Sra. Towards an optimal stochastic alternating direction method of multipliers. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*. Vol. 32. JMLR Workshop and Conference Proceedings, pages 620–628, 2014 (cited on page 16).

[367] M. Kiefel, C. Schuler, P. Hennig. Probabilistic Progress Bars. In *Conference on Pattern Recognition (GCPR)*. Vol. 8753. Lecture Notes in Computer Science, pages 331–341, 2014.

[368] M. Kiefel, P. Gehler. Human Pose Estimation with Fields of Parts. In *13th European Conference on Computer Vision (ECCV 2014)*. Vol. 8693. Lecture Notes in Computer Science, pages 331–346, 2014.

[369] P. Hennig, S. Hauberg. Probabilistic Solutions to Differential Equations and their Application to Riemannian Statistics. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS 2014)*. Vol. 33. JMLR Workshop and Conference Proceedings, pages 347–355, 2014 (cited on page 22).

[370] K. Muandet, K. Fukumizu, B. Sriperumbudur, A. Gretton, B. Schölkopf. Kernel Mean Estimation and Stein Effect. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*. Vol. 32. JMLR Workshop and Conference Proceedings, pages 10–18, 2014 (cited on page 14).

[371] R. Garnett, M. Osborne, P. Hennig. Active Learning of Linear Embeddings for Gaussian Processes. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)*, pages 230–239, 2014 (cited on page 7).

[372] M. Sugiyama, C.-A. Azencott, D. Grimm, Y. Kawahara, K. Borgwardt. Multi-Task Feature Selection on Multiple Networks via Maximum Flows. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 199–207, 2014.

[373] D. Lopez-Paz, S. Sra, A. Smola, Z. Ghahramani, B. Schölkopf. Randomized Nonlinear Component Analysis. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*. Vol. 32. JMLR Workshop and Conference Proceedings, pages 1359–1367, 2014.

[374] M. Gomez Rodriguez, K. Gummadi, B. Schölkopf. Quantifying Information Overload in Social Media and its Impact on Social Contagions. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM 2014)*, pages 170–179, 2014.

[375] H. Daneshmand, M. Gomez Rodriguez, L. Song, B. Schölkopf. Estimating Diffusion Network Structures: Recovery Conditions, Sample Complexity & Soft-thresholding Algorithm. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*. Vol. 32. JMLR Workshop and Conference Proceedings, pages 793–801, 2014 (cited on page 17).

[376] S. Weichwald, B. Schölkopf, T. Ball, M. Grosse-Wentrup. Causal and Anti-Causal Learning in Pattern Recognition for Neuroimaging. In *4th International Workshop on Pattern Recognition in Neuroimaging (PRNI 2014)*, 2014.

[377] S. Weichwald, T. Meyer, B. Schölkopf, T. Ball, M. Grosse-Wentrup. Decoding Index Finger Position from EEG Using Random Forests. In *4th International Workshop on Cognitive Information Processing (CIP 2014)*, 2014.

[378] B. Bischoff, D. Nguyen-Tuong, H. van Hoof, A. McHutchon, C. Rasmussen, A. Knoll, J. Peters, M. Deisenroth. Policy Search For Learning Robot Control Using Sparse Data. In *Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA 2014)*, pages 3882–3887, 2014.

[379] H. Ben Amor, G. Neumann, S. Kamthe, O. Kroemer, J. Peters. Interaction Primitives for Human-Robot Cooperation Tasks. In *Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA 2014)*, pages 2831–2837, 2014.

[380] R. Calandra, N. Gopalan, A. Seyfarth, J. Peters, M. Deisenroth. Bayesian Gait Optimization for Bipedal Locomotion. In *Proceedings of the 8th International Conference on Learning and Intelligent Optimization (LION 2014)*. Vol. 8426. Lecture Notes in Computer Science, pages 274–290, 2014.

[381] R. Calandra, A. Seyfarth, J. Peters, M. Deisenroth. An Experimental Comparison of Bayesian Optimization for Bipedal Locomotion. In *Proceedings of 2014 IEEE International Conference on Robotics and Automation*, pages 1951–1958, 2014.

[382] C. Daniel, M. Viering, J. Metz, O. Kroemer, J. Peters. Active Reward Learning. In *Proceedings of Robotics: Science and Systems X*, 2014.

[383] M. Deisenroth, P. Englert, J. Peters, D. Fox. Multi-Task Policy Search for Robotics. In *Proceedings of 2014 IEEE International Conference on Robotics and Automation*, pages 3876–3881, 2014.

[384] S. Kamthe, J. Peters, M. Deisenroth. Multi-modal filtering for non-linear estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pages 7979–7983, 2014.

[385] O. Kroemer, H. van Hoof, G. Neumann, J. Peters. Learning to Predict Phases of Manipulation Tasks as Hidden States (ICRA 2014). In *Proceedings of 2014 IEEE International Conference on Robotics and Automation*, pages 4009–4014, 2014.

[386] R. Lioutikov, O. Kroemer, J. Peters, G. Maeda. Learning Manipulation by Sequencing Motor Primitives with a Two-Armed Robot. In *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS 2014)*. Vol. 302. Advances in Intelligent Systems and Computing, pages 1601–1611, 2014.

[387] R. Lioutikov, A. Paraschos, J. Peters, G. Neumann. Sample-Based Information-Theoretic Stochastic Optimal Control. In *Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA 2014)*, pages 3896–3902, 2014.

[388] V. Wiens, L. Schlaffke, T. Schmidt-Wilcke, T. Schultz. Visualizing Uncertainty in HARDI Tractography Using Superquadric Streamtubes. In *Eurographics Conference on Visualization, Short Papers (Eurovis 2014)*, 2014 (cited on page 15).

[389] M. Hermann, A. Schunke, T. Schultz, R. Klein. A Visual Analytics Approach to Study Anatomic Covariation. In *Proceedings of IEEE Pacific Visualization 2014*, pages 161–168, 2014.

[390] G. Doran, K. Muandet, K. Zhang, B. Schölkopf. A Permutation-Based Kernel Conditional Independence Test. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)*, pages 132–141, 2014 (cited on pages 5, 7, 15).

[391] O. Özdenizci, T. Meyer, M. Cetin, M. Grosse-Wentrup. Towards Neurofeedback Training of Associative Brain Areas for Stroke Rehabilitation. In *6th International Brain-Computer Interface Conference (BCI 2014)*, 2014.

[392] R. Chaves, L. Luft, T. Maciel, D. Gross, D. Janzing, B. Schölkopf. Inferring latent structures via information inequalities. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)*, pages 112–121, 2014.

[393] A. Argyriou, F. Dinuzzo. A unifying view of representer theorems. In *Proceedings of the 31th International Conference on Machine Learning (ICML 2014)*. Vol. 32. JMLR Workshop and Conference Proceedings, pages 748–756, 2014.

[394] A. Cherian, S. Sra. Riemannian Sparse Coding for Positive Definite Matrices. In *13th European Conference on Computer Vision*. Vol. LNCS 8691, pages 299–314, 2014.

[395] K. Muandet, B. Sriperumbudur, B. Schölkopf. Kernel Mean Estimation via Spectral Filtering. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pages 1–9, 2014 (cited on pages 5, 14).

[396] T. Gunter, M. Osborne, R. Garnett, P. Hennig, S. Roberts. Sampling for Inference in Probabilistic Models with Fast Bayesian Quadrature. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pages 2789–2797, 2014 (cited on pages 7, 22).

[397] M. Schober, D. Duvenaud, P. Hennig. Probabilistic ODE Solvers with Runge-Kutta Means. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pages 739–747, 2014 (cited on page 21).

[398] M. Farajtabar, N. Du, M. Gomez Rodriguez, I. Valera, H. Zha, L. Song. Shaping Social Activity by Incentivizing Users. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 2014.

[399] M. Schober, N. Kasenburg, A. Feragen, P. Hennig, S. Hauberg. Probabilistic Shortest Path Tractography in DTI Using Gaussian Process ODE Solvers. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014, Lecture Notes in Computer Science Vol. 8675*, pages 265–272, 2014 (cited on page 22).

[400] S. Kpotufe, E. Sgouritsa, D. Janzing, B. Schölkopf. Consistency of Causal Inference under the Additive Noise Model. In *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*. Vol. 32. JMLR Workshop and Conference Proceedings, pages 478–495, 2014 (cited on pages 7, 12).

[401] P. Geiger, D. Janzing, B. Schölkopf. Estimating Causal Effects by Bounding Confounding. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)*, pages 240–249, 2014 (cited on page 18).

[402] R. Köhler, C. Schuler, B. Schölkopf, S. Harmeling. Mask-Specific Inpainting with Deep Neural Networks. In *Pattern Recognition (GCPR 2014)*, pages 523–534, 2014 (cited on page 24).

[403] F. Meier, P. Hennig, S. Schaal. Efficient Bayesian Local Model Learning for Control. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS 2014)*, pages 2244–2249, 2014.

[404] F. Meier, P. Hennig, S. Schaal. Incremental Local Gaussian Regression. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pages 972–980, 2014 (cited on page 7).

2013

[405] D. Balduzzi. Falsification and future performance. In *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence (Solomonoff 85th Memorial Conference)*. Vol. 7070. Lecture Notes in Computer Science, pages 65–78, 2013.

[406] P. Hennig. Fast Probabilistic Optimization from Noisy Gradients. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*. Vol. 28(1). JMLR Workshop and Conference Proceedings, pages 62–70, 2013 (cited on page 21).

[407] K. Muandet, D. Balduzzi, B. Schölkopf. Domain Generalization via Invariant Feature Representation. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*. Vol. 28(1). JMLR Workshop and Conference Proceedings, pages 10–18, 2013 (cited on page 14).

[408] M. Gomez Rodriguez, J. Leskovec, B. Schölkopf. Structure and Dynamics of Information Pathways in On-line Media. In *6th ACM International Conference on Web Search and Data Mining (WSDM 2013)*, pages 23–32, 2013.

[409] C. Daniel, G. Neumann, J. Peters. Learning Sequential Motor Tasks. In *Proceedings of 2013 IEEE International Conference on Robotics and Automation (ICRA 2013)*, 2013.

[410] P. Englert, A. Paraschos, J. Peters, M. Deisenroth. Model-based Imitation Learning by Probabilistic Trajectory Matching. In *Proceedings of 2013 IEEE International Conference on Robotics and Automation (ICRA 2013)*, pages 1922–1927, 2013.

[411] N. Gopalan, M. Deisenroth, J. Peters. Feedback Error Learning for Rhythmic Motor Primitives. In *Proceedings of 2013 IEEE International Conference on Robotics and Automation (ICRA 2013)*, pages 1317–1322, 2013.

[412] Y. Seldin, C. Szepesvári, P. Auer, Y. Abbasi-Yadkori. Evaluation and Analysis of the Performance of the EXP3 Algorithm in Stochastic Environments. In *Proceedings of the Tenth European Workshop on Reinforcement Learning (EWRL 2012)*. Vol. 24. JMLR Workshop and Conference Proceedings, pages 103–116, 2012.

[413] D. Lopez-Paz, J. Hernandez-Lobato, Z. Ghahramani. Gaussian Process Vine Copulas for Multivariate Dependence. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*. Vol. 28(2). JMLR Workshop and Conference Proceedings, pages 10–18, 2013.

[414] A. Kupcsik, M. Deisenroth, J. Peters, G. Neumann. Data-Efficient Generalization of Robot Skills with Contextual Policy Search. In *Proceedings of the 27th National Conference on Artificial Intelligence (AAAI 2013)*, 2013.

[415] K. Muandet, B. Schölkopf. One-class Support Measure Machines for Group Anomaly Detection. In *Proceedings 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013)*, pages 449–458, 2013 (cited on pages 5, 14).

[416] T. Zander, B. Battes, B. Schölkopf, M. Grosse-Wentrup. Towards neurofeedback for improving visual attention. In *Proceedings of the Fifth International Brain-Computer Interface Meeting: Defining the Future*, Article ID: 086, 2013.

[417] K. Zhang, B. Schölkopf, K. Muandet, Z. Wang. Domain adaptation under Target and Conditional Shift. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*. Vol. 28(3. JMLR Workshop and Conference Proceedings, pages 819–827, 2013.

[418] M. Gomez Rodriguez, J. Leskovec, B. Schölkopf. Modeling Information Propagation with Survival Theory. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*. Vol. 28(3). JMLR Workshop and Conference Proceedings, pages 666–674, 2013 (cited on page 17).

[419] D. Lopez-Paz, P. Hennig, B. Schölkopf. The Randomized Dependence Coefficient. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1–9, 2013.

[420] M. Grosse-Wentrup, S. Harmeling, T. Zander, J. Hill, B. Schölkopf. How to Test the Quality of Reconstructed Sources in Independent Component Analysis (ICA) of EEG/MEG Data. In *Proceedings of the 3rd International Workshop on Pattern Recognition in NeuroImaging (PRNI)*, pages 102–105, 2013.

[421] J. Mooij, D. Janzing, B. Schölkopf. From Ordinary Differential Equations to Structural Causal Models: the deterministic case. In *Proceedings 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013)*, pages 440–448, 2013.

[422] E. Sgouritsa, D. Janzing, J. Peters, B. Schölkopf. Identifying Finite Mixtures of Nonparametric Product Distributions and Causal Inference of Confounders. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013)*, pages 556–565, 2013 (cited on pages 15, 18).

[423] S. Harmeling, M. Hirsch, B. Schölkopf. On a link between kernel mean maps and Fraunhofer diffraction, with an application to super-resolution beyond the diffraction limit. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013)*, pages 1083–1090, 2013 (cited on pages 5, 23).

[424] C. Schuler, H. Burger, S. Harmeling, B. Schölkopf. A machine learning approach for non-blind image deconvolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013)*, pages 1067–1074, 2013 (cited on page 24).

[425] R. Köhler, M. Hirsch, B. Schölkopf, S. Harmeling. Improving alpha matting and motion blurred foreground estimation. In *IEEE Conference on Image Processing (ICIP 2013)*, pages 3446–3450, 2013.

[426] H. Ben Amor, D. Vogt, M. Ewerton, E. Berger, B. Jung, J. Peters. Learning responsive robot behavior by imitation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2013)*, pages 3257–3264, 2013.

[427] J. Peters, J. Kober, K. Mülling, O. Kroemer, G. Neumann. Towards Robot Skill Learning: From Simple Skills to Table Tennis. In *Machine Learning and Knowledge Discovery in Databases, Proceedings of the European Conference on Machine Learning, Part III (ECML 2013)*. Vol. LNCS 8190, pages 627–631, 2013.

[428] G. Neumann, A. Kupcsik, M. Deisenroth, J. Peters. Information-Theoretic Motor Skill Learning. In *Proceedings of the 27th AAAI 2013, Workshop on Intelligent Robotic Systems (AAAI 2013)*, 2013.

[429] N. Du, L. Song, M. Gomez Rodriguez, H. Zha. Scalable Influence Estimation in Continuous-Time Diffusion Networks. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 3147–3155, 2013 (cited on page 17).

[430] Z. Chen, K. Zhang, L. Chan. Nonlinear Causal Discovery for High Dimensional Data: A Kernelized Trace Method. In *IEEE 13th International Conference on Data Mining (ICDM 2013)*, pages 1003–1008, 2013.

[431] M. Sugiyama, K. Borgwardt. Measuring Statistical Dependence via the Mutual Information Dimension. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, pages 1692–1698, 2013.

[432] M. Sugiyama, K. Borgwardt. Rapid Distance-Based Outlier Detection via Sampling. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 467–475, 2013.

[433] A. Paraschos, C. Daniel, J. Peters, G. Neumann. Probabilistic Movement Primitives. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2616–2624, 2013.

[434] A. Paraschos, G. Neumann, J. Peters. A probabilistic approach to robot trajectory generation. In *Proceedings of the 13th IEEE International Conference on Humanoid Robots (HUMANOIDS)*, pages 477–483, 2013.

[435] H. van Hoof, O. Kroemer, J. Peters. Object Modeling and Segmentation by Robot Interaction with Cluttered Environments. In *Proceedings of the 13th IEEE International Conference on Humanoid Robots (HUMANOIDS)*, pages 169–176, 2013.

[436] E. Klenske, M. Zeilinger, B. Schölkopf, P. Hennig. Nonparametric dynamics estimation for time periodic systems. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*, pages 486–493, 2013.

[437] M. Bangert, P. Hennig, U. Oelfke. Analytical probabilistic proton dose calculation and range uncertainties. In *17th International Conference on the Use of Computers in Radiation Therapy*, pages 6–11, 2013 (cited on page 22).

[438] J. Peters, D. Janzing, B. Schölkopf. Causal Inference on Time Series using Restricted Structural Equation Models. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 154–162, 2013 (cited on page 18).

[439] S. Kpotufe, V. Garg. Adaptivity to Local Smoothness and Dimension in Kernel Regression. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 3075–3083, 2013 (cited on page 12).

[440] S. Kpotufe, F. Orabona. Regression-tree Tuning in a Streaming Setting. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1788–1796, 2013 (cited on page 12).

[441] M. Besserve, N. Logothetis, B. Schölkopf. Statistical analysis of coupled time series with Kernel Cross-Spectral Density operators. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2535–2543, 2013 (cited on pages 14, 31).

[442] U. von Luxburg, M. Alamgir. Density estimation from unweighted k-nearest neighbor graphs: a roadmap. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 225–233, 2013.

[443] S. Sra, R. Hosseini. Geometric optimisation on positive definite matrices for elliptically contoured distributions. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2562–2570, 2013 (cited on page 16).

[444] S. Jegelka, F. Bach, S. Sra. Reflection methods for user-friendly submodular optimization. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1313–1321, 2013 (cited on page 16).

[445] A. Feragen, N. Kasenburg, J. Petersen, M. de Bruijne, K. Borgwardt. Scalable kernels for graphs with continuous attributes. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 216–224, 2013.

[446] B. Rakitsch, C. Lippert, K. Borgwardt, O. Stegle. It is all in the noise: Efficient multi-task Gaussian process inference with structured residuals. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1466–1474, 2013.

[447] T. Schultz, S. Gröschel. Auto-Calibrating Spherical Deconvolution Based on ODF Sparsity. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention, Part I (MICCAI 2013)*, pages 663–670, 2013.

2012

[448] P. Hennig, M. Kiefel. Quasi-Newton Methods: A New Direction. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 25–32, 2012 (cited on page 21).

[449] U. von Luxburg, R. Williamson, I. Guyon. Clustering: Science or Art? In *Proceedings of the Unsupervised and Transfer Learning Workshop at ICML 2011*. Vol. 27. JMLR Workshop and Conference Proceedings, pages 65–79, 2012.

[450] H. Burger, C. Schuler, S. Harmeling. Image denoising: Can plain Neural Networks compete with BM3D? In *25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, pages 2392–2399, 2012 (cited on pages 7, 24).

[451] O. Stegle, F. Roth, Q. Morris, J. Listgarten. Personalized medicine: from genotypes and molecular phenotypes towards computed therapy. In *Pacific Symposium on Biocomputing (PSB 2012)*, pages 323–326, 2012.

[452] P. Hennig, D. Stern, R. Herbrich, T. Graepel. Kernel Topic Models. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012)*. Vol. 22. JMLR Workshop and Conference Proceedings, pages 511–519, 2012 (cited on page 7).

[453] B. Bócsi, P. Hennig, L. Csató, J. Peters. Learning Tracking Control with Forward Models. In *IEEE International Conference on Robotics and Automation (ICRA 2012)*, pages 259–264, 2012.

[454] Y. Seldin, N. Cesa-Bianchi, P. Auer, F. Laviolette, J. Shawe-Taylor. PAC-Bayes-Bernstein Inequality for Martingales and its Application to Multiarmed Bandits. In *Proceedings of the Workshop on On-line Trading of Exploration and Exploitation 2 at ICML 2011*. Vol. 26. JMLR Workshop and Conference Proceedings, pages 98–111, 2012.

[455] C. Daniel, G. Neumann, J. Peters. Hierarchical Relative Entropy Policy Search. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012)*. Vol. 22. JMLR Workshop and Conference Proceedings, pages 273–281, 2012 (cited on pages 8, 9, 28).

[456] O. Kroemer, E. Ugur, E. Oztop, J. Peters. A Kernel-based Approach to Direct Action Perception. In *International Conference on Robotics and Automation (ICRA 2012)*, pages 2605–2610, 2012.

[457] M. Gomez Rodriguez, B. Schölkopf. Submodular Inference of Diffusion Networks from Multiple Trees. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 489–496, 2012 (cited on page 17).

[458] M. Gomez Rodriguez, B. Schölkopf. Influence Maximization in Continuous Time Diffusion Networks. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 313–320, 2012 (cited on page 17).

[459] Z. Wang, M. Deisenroth, H. Ben Amor, D. Vogt, B. Schölkopf, J. Peters. Probabilistic Modeling of Human Movements for Intention Inference. In *Proceedings of Robotics: Science and Systems VIII*, pages 1–8, 2012 (cited on page 27).

[460] T. Meyer, J. Peters, D. Brötz, T. Zander, B. Schölkopf, S. Soekadar, M. Grosse-Wentrup. A Brain-Robot Interface for Studying Motor Learning after Stroke. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 4078–4083, 2012.

[461] M. Alamgir, U. von Luxburg. Shortest path distance in random k-nearest neighbor graphs. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, 2012 (cited on page 13).

[462] Y. Nishiyama, A. Boularias, A. Gretton, K. Fukumizu. Hilbert Space Embeddings of POMDPs. In *Conference on Uncertainty in Artificial Intelligence (UAI 2012)*, 2012 (cited on pages 5, 15).

[463] A. Boularias, O. Kroemer, J. Peters. Structured Apprenticeship Learning. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2012)*, 2012 (cited on pages 9, 29).

[464] H. Ben Amor, O. Kroemer, U. Hillenbrand, G. Neumann, J. Peters. Generalization of Human Grasping for Multi-Fingered Robot Hands. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 2043–2050, 2012.

[465] C. Daniel, G. Neumann, J. Peters. Learning Concurrent Motor Skills in Versatile Solution Spaces. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 3591–3597, 2012 (cited on pages 8, 9, 27).

[466] M. Deisenroth, R. Calandra, A. Seyfarth, J. Peters. Toward Fast Policy Search for Learning Legged Locomotion. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 1787–1792, 2012.

[467] J. Kober, K. Mülling, J. Peters. Learning Throwing and Catching Skills. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 5167–5168, 2012 (cited on page 8).

[468] J. Peters, J. Kober, K. Mülling, D. Nguyen-Tuong, O. Kroemer. Robot Skill Learning. In *20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 40–45, 2012 (cited on page 27).

[469] H. van Hoof, O. Kroemer, H. Ben Amor, J. Peters. Maximally Informative Interaction Learning for Scene Exploration. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 5152–5158, 2012.

[470] B. Schölkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, J. Mooij. On Causal and Anticausal Learning. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 1255–1262, 2012 (cited on pages 6, 7, 19).

[471] T. Meyer, J. Peters, T. Zander, D. Brötz, S. Soekadar, B. Schölkopf, M. Grosse-Wentrup. Investigating the Neural Basis of Brain-Computer Interface (BCI)-based Stroke Rehabilitation. In *International Conference on NeuroRehabilitation (ICNR 2012)*, pages 617–621, 2012.

[472] C. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf. Blind Correction of Optical Aberrations. In *12th European Conference on Computer Vision - ECCV 2012, LNCS Vol. 7574*, pages 187–200, 2012 (cited on pages 8, 23, 24).

[473] D. Balduzzi, M. Besserve. Towards a learning-theoretic analysis of spike-timing dependent plasticity. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 2465–2473, 2012 (cited on pages 10, 31).

[474] P. Ortega, J. Grau-Moya, T. Genewein, D. Balduzzi, D. Braun. A Nonparametric Conjugate Prior Distribution for the Maximizing Argument of a Noisy Function. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 3014–3022, 2012.

[475] C. Persello, F. Dinuzzo. Interactive Domain Adaptation Technique for the Classification of Remote Sensing Images. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2012)*, pages 6872–6875, 2012.

[476] O. Kroemer, H. Ben Amor, M. Ewerton, J. Peters. Point Cloud Completion Using Symmetries and Extrusions. In *IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS 2012)*, pages 680–685, 2012.

[477] A. Boularias, O. Kroemer, J. Peters. Algorithms for Learning Markov Field Policies. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 2186–2194, 2012.

[478] K. Muandet, K. Fukumizu, F. Dinuzzo, B. Schölkopf. Learning from distributions via support measure machines. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 10–18, 2012 (cited on pages 5, 14).

[479] D. Lopez-Paz, J. Hernandez-Lobato, B. Schölkopf. Semi-Supervised Domain Adaptation with Copulas. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 674–682, 2012.

[480] Z. Chen, K. Zhang, L. Chan. Causal discovery with scale-mixture model for spatiotemporal variance dependencies. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 1736–1744, 2012.

[481] S. Kpotufe, A. Boularias. Gradient Weights help Nonparametric Regressors. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 2870–2878, 2012 (cited on pages 4, 12).

[482] M. Gomez Rodriguez, M. Rogati. Bridging Offline and Online Social Graph Dynamics. In *21st ACM Conference on Information and Knowledge Management (CIKM 2012)*, pages 2447–2450, 2012.

[483] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, S. Harmeling. Recording and Playback of Camera Shake: Benchmarking Blind Deconvolution with a Real-World Database. In *12th European Conference on Computer Vision - ECCV 2012, LNCS Vol. 7578*, pages 27–40, 2012 (cited on page 23).

[484] M. Hirsch, M. Hofmann, F. Mantlik, B. Pichler, B. Schölkopf, M. Habeck. A Blind Deconvolution Approach for Pseudo CT Prediction from MR Image Pairs. In *19th IEEE International Conference on Image Processing (ICIP 2012)*, pages 2953–2956, 2012.

[485] S. Sra. Scalable nonconvex inexact proximal splitting. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 539–547, 2012.

[486] S. Sra. A new metric on the manifold of kernel matrices with application to matrix geometric means. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 144–152, 2012.

[487] F. Dinuzzo, B. Schölkopf. The representer theorem for Hilbert spaces: a necessary and sufficient condition. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 189–196, 2012 (cited on page 5).

[488] M. Habeck. Evaluation of marginal likelihoods via the density of states. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012)*. Vol. 22. JMLR Workshop and Conference Proceedings, pages 486–494, 2012.

[489] T. Zander, A. Pape. Towards identifying and validating cognitive correlates in a passive Brain-Computer Interface for detecting Loss of Control. In *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2012)*, 2012.

[490] A. Tawari, C. Tran, A. Doshi, T. Zander. Distributed multisensory signals acquisition and analysis in dyadic interactions. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems Extended Abstracts*, pages 2261–2266, 2012.

[491] S. Grünewälder, G. Lever, L. Baldassarre, M. Pontil, A. Gretton. Modelling transition dynamics in MDPs with RKHS embeddings. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 535–542, 2012 (cited on page 15).

[492] S. Grünewälder, G. Lever, L. Baldassarre, S. Patterson, A. Gretton, M. Pontil. Conditional mean embeddings as regressors. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 1823–1830, 2012 (cited on page 15).

[493] D. Sejdinovic, A. Gretton, B. Sriperumbudur, K. Fukumizu. Hypothesis testing using pairwise distances and associated kernels. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, pages 1111–1118, 2012.

[494] A. Gretton, B. Sriperumbudur, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, K. Fukumizu. Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 1214–1222, 2012 (cited on page 14).

[495] C. Widmer, M. Kloft, N. Görnitz, G. Rätsch. Efficient Training of Graph-Regularized Multitask SVMs. In *Machine Learning and Knowledge Discovery in Databases - European Conference (ECML/PKDD 2012)*. Vol. LNCS 7523, pages 633–647, 2012.

2011

[496] P. Gehler, C. Rother, M. Kiefel, L. Zhang, B. Schölkopf. Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 765–773, 2011.

[497] M. Alamgir, U. von Luxburg. Phase transition in the family of p-resistances. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 379–387, 2011 (cited on page 13).

[498] M. Seeger, H. Nickisch. Fast Convergent Algorithms for Expectation Propagation Approximate Bayesian Inference. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*. Vol. 15. JMLR Workshop and Conference Proceedings, pages 652–660, 2011 (cited on page 7).

[499] J. Kober, E. Oztop, J. Peters. Reinforcement Learning to adjust Robot Movements to New Situations. In *Proceedings of Robotics: Science and Systems VI*, pages 33–40, 2011 (cited on page 8).

[500] M. Besserve, D. Janzing, N. Logothetis, B. Schölkopf. Finding dependencies between frequencies with the kernel cross-spectral density. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, pages 2080–2083, 2011.

[501] O. Kroemer, J. Peters. A Flexible Hybrid Framework for Modeling Complex Manipulation Tasks. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2011)*, pages 1856–1861, 2011.

[502] R. Lampariello, D. Nguyen-Tuong, C. Castellini, G. Hirzinger, J. Peters. Trajectory Planning for Optimal Robot Catching in Real-Time. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2011)*, pages 3719–3726, 2011.

[503] J. Peters, K. Mülling, J. Kober, D. Nguyen-Tuong, O. Kroemer. Towards Motor Skill Learning for Robotics. In *Robotics Research*, pages 469–482, 2011.

[504] J. Piater, S. Jodogne, R. Detry, D. Kraft, N. Krüger, O. Kroemer, J. Peters. Learning Visual Representations for Interactive Systems. In *Robotics Research*, pages 399–416, 2011.

[505] S. Jegelka, J. Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pages 1897–1904, 2011.

[506] M. Grosse-Wentrup. Fronto-Parietal Gamma-Oscillations are a Cause of Performance Variation in Brain-Computer Interfacing. In *5th International IEEE/EMBS Conference on Neural Engineering (NER 2011)*, pages 384–387, 2011.

[507] M. Langovoy, M. Habeck, B. Schölkopf. Adaptive nonparametric detection in cryo-electron microscopy. In *Proceedings of the 58th World Statistics Congress*, pages 4456–4461, 2011.

[508] A. Boularias, J. Kober, J. Peters. Relative Entropy Inverse Reinforcement Learning. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*. Vol. 15. JMLR Workshop and Conference Proceedings, pages 182–189, 2011 (cited on pages 9, 29).

[509] J. Kober, E. Oztop, J. Peters. Reinforcement Learning to adjust Robot Movements to New Situations. In *Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 2650–2655, 2011.

[510] M. Gomez Rodriguez, M. Grosse-Wentrup, J. Hill, A. Gharabaghi, B. Schölkopf, J. Peters. Towards Brain-Robot Interfaces in Stroke Rehabilitation. In *12th International Conference on Rehabilitation Robotics (ICORR 2011)*, page 6, 2011.

[511] J. Peters, J. Mooij, D. Janzing, B. Schölkopf. Identifiability of causal graphs using functional models. In *27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, pages 589–598, 2011.

[512] M. Gomez Rodriguez, D. Balduzzi, B. Schölkopf. Uncovering the Temporal Dynamics of Diffusion Networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 561–568, 2011 (cited on page 17).

[513] P. Achlioptas, B. Schölkopf, K. Borgwardt. Two-locus association mapping in subquadratic time. In *17th ACM SIGKKD Conference on Knowledge Discovery and Data Mining (KDD 2011)*, pages 726–734, 2011 (cited on page 10).

[514] S. Jegelka, J. Bilmes. Approximation Bounds for Inference using Cooperative Cut. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 577–584, 2011.

[515] S. Jegelka, J. Bilmes. Online submodular minimization for combinatorial structures. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 345–352, 2011.

[516] S. Kpotufe, U. von Luxburg. Pruning nearest neighbor cluster trees. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 225–232, 2011.

[517] D. García-García, U. von Luxburg, R. Santos-Rodríguez. Risk-Based Generalizations of f-divergences. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 417–424, 2011.

[518] J. Zscheischler, D. Janzing, K. Zhang. Testing whether linear equations are causal: A free probability theory approach. In *27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, pages 839–847, 2011 (cited on page 18).

[519] K. Zhang, J. Peters, D. Janzing, B. Schölkopf. Kernel-based Conditional Independence Test and Application in Causal Discovery. In *27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, pages 804–813, 2011 (cited on pages 5, 7, 15).

[520] M. Grosse-Wentrup. Neurofeedback of Fronto-Parietal Gamma-Oscillations. In *Proceedings of the 5th International Brain-Computer Interface Conference (BCI 2011)*, pages 172–175, 2011.

[521] J. Saab, B. Battes, M. Grosse-Wentrup. Simultaneous EEG Recordings with Dry and Wet Electrodes in Motor-Imagery. In *Proceedings of the 5th International Brain-Computer Interface Conference (BCI 2011)*, pages 312–315, 2011.

[522] K. Zhang, A. Hyvärinen. A general linear non-Gaussian state-space model: Identifiability, identification, and applications. In *Proceedings of the 3rd Asian Conference on Machine Learning (ACML 2011)*. Vol. 20. JMLR Workshop and Conference Proceedings, pages 113–128, 2011.

[523] S. Jegelka, H. Lin, J. Bilmes. On Fast Approximate Submodular Minimization. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 460–468, 2011.

[524] D. Janzing, E. Sgouritsa, O. Stegle, J. Peters, B. Schölkopf. Detecting low-complexity unobserved causes. In *27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, pages 383–391, 2011.

[525] Z. Wang, A. Boularias, K. Mülling, J. Peters. Balancing Safety and Exploitability in Opponent Modeling. In *Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2011)*, pages 1515–1520, 2011.

[526] Z. Wang, C. Lampert, K. Mülling, B. Schölkopf, J. Peters. Learning anticipation policies for robot table tennis. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, pages 332–337, 2011.

[527] D. Balduzzi. Detecting emergent processes in cellular automata with excess information. In *Advances in Artificial Life: ECAL 2011*, pages 55–62, 2011.

[528] D. Balduzzi. Estimating integrated information with TMS pulses during wakefulness, sleep and under anesthesia. In *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2011)*, pages 4717–4720, 2011.

[529] F. Dinuzzo, K. Fukumizu. Learning low-rank output kernels. In *Proceedings of the 3rd Asian Conference on Machine Learning (ACML 2011)*. Vol. 20. JMLR Workshop and Conference Proceedings, pages 181–196, 2011.

[530] O. Kroemer, J. Peters. A Non-Parametric Approach to Dynamic Programming. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 1719–1727, 2011 (cited on page 28).

[531] A. Boularias, O. Kroemer, J. Peters. Learning robot grasping from 3-D images with Markov Random Fields. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, pages 1548–1553, 2011.

[532] B. Bocsi, D. Nguyen-Tuong, L. Csato, B. Schölkopf, J. Peters. Learning inverse kinematics with structured prediction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, pages 698–703, 2011.

[533] D. Nguyen-Tuong, J. Peters. Learning task-space tracking control with kernels. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, pages 704–709, 2011.

[534] J. Kober, J. Peters. Learning elementary movements jointly with a higher level task. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, pages 338–343, 2011.

[535] C. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf. Non-stationary correction of optical aberrations. In *13th IEEE International Conference on Computer Vision (ICCV 2011)*, pages 659–666, 2011 (cited on pages 8, 23, 24).

[536] Y. Seldin, P. Auer, F. Laviolette, J. Shawe-Taylor, R. Ortner. PAC-Bayesian Analysis of Contextual Bandits. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 1683–1691, 2011 (cited on pages 5, 9, 12, 29).

[537] M. Hirsch, C. Schuler, S. Harmeling, B. Schölkopf. Fast removal of non-uniform camera shake. In *13th IEEE International Conference on Computer Vision (ICCV 2011)*, pages 463–470, 2011 (cited on pages 7, 23, 24).

[538] H. Burger, B. Schölkopf, S. Harmeling. Removing noise from astronomical images using a pixel-specific noise model. In *IEEE International Conference on Computational Photography (ICCP 2011)*, page 8, 2011 (cited on page 24).

[539] H. Burger, S. Harmeling. Improving Denoising Algorithms via a Multi-scale Meta-procedure. In *Pattern Recognition, 33rd DAGM Symposium*, pages 206–215, 2011 (cited on pages 7, 24).

[540] P. Hennig. Optimal Reinforcement Learning for Gaussian Systems. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 325–333, 2011 (cited on pages 9, 29).

[541] J. Mooij, D. Janzing, B. Schölkopf, T. Heskes. On Causal Discovery with Cyclic Additive Noise Models. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 639–647, 2011.

[542] V. Franc, A. Zien, B. Schölkopf. Support Vector Machines as Probabilistic Models. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 665–672, 2011.

[543] D. Duvenaud, H. Nickisch, C. Rasmussen. Additive Gaussian Processes. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 226–234, 2011 (cited on page 7).

[544] A. Loktyushin, S. Harmeling. Automatic foreground-background refocusing. In *18th IEEE International Conference on Image Processing (ICIP 2011)*, pages 3445–3448, 2011.

[545] P. Joubert, S. Nickell, F. Beck, M. Habeck, M. Hirsch, B. Schölkopf. Automatic particle picking using diffusion filtering and random forest classification. In *International Workshop on Microscopic Image Analysis with Application in Biology (MIAAB 2011)*, pages 1–6, 2011.

[546] S. Kpotufe. k-NN Regression Adapts to Local Intrinsic Dimension. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 729–737, 2011 (cited on pages 4, 12).

[547] O. Stegle, C. Lippert, J. Mooij, N. Lawrence, K. Borgwardt. Efficient inference in matrix-variate Gaussian models with iid observation noise. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 630–638, 2011.

[548] A. Cherian, S. Sra, A. Banerjee, N. Papanikolopoulos. Efficient Similarity Search for Covariance Matrices via the Jensen-Bregman LogDet Divergence. In *IEEE International Conference on Computer Vision (ICCV 2011)*, pages 2399–2406, 2011.

[549] S. Sra, A. Cherian. Generalized Dictionary Learning for Symmetric Positive Definite Matrices with Application to Nearest Neighbor Retrieval. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2011, LNCS vol 6913, Part III*, pages 318–332, 2011.

[550] S. Sra. Fast projections onto L1,q-norm balls for grouped feature selection. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2011, LNCS vol 6913, Part III*, pages 305–317, 2011.

[551] A. Barbero, S. Sra. Fast Newton-type Methods for Total-Variation with Applications. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 313–320, 2011.

[552] A. Cherian, S. Sra, N. Papanikolopoulos. Denoising sparse noise via online dictionary learning. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, pages 2060–2063, 2011.

[553] T. Zander, D. Klippel, R. Scherer. Introducing the detection of auditory error responses based on BCI technology for passive interaction. In *Proceedings of the 5th International Brain-Computer Interface Conference (BCI 2011)*, pages 252–255, 2011.

[554] D. Balderas, T. Zander, F. Bachl, C. Neuper, R. Scherer. Restricted boltzmann machines as useful tool for detecting oscillatory eeg components. In *Proceedings of the 5th International Brain-Computer Interface Conference (BCI 2011)*, pages 68–71, 2011.

[555] M. Deisenroth, C. Rasmussen. PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pages 465–472, 2011 (cited on page 29).

[556] L. Song, A. Gretton, D. Bickson, Y. Low, C. Guestrin. Kernel Belief Propagation. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*. Vol. 15. JMLR Workshop and Conference Proceedings, pages 707–715, 2011 (cited on pages 5, 15).

[557] J. Gonzalez, Y. Low, A. Gretton, C. Guestrin. Parallel Gibbs Sampling: From Colored Fields to Thin Junction Trees. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*. Vol. 15. JMLR Workshop and Conference Proceedings, pages 324–332, 2011.

[558] K. Fukumizu, L. Song, A. Gretton. Kernel Bayes' Rule. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 1737–1745, 2011 (cited on pages 5, 15).

[559] N. Görnitz, C. Widmer, G. Zeller, A. Kahles, S. Sonnenburg, G. Rätsch. Hierarchical Multitask Structured Output Learning for Large-scale Sequence Segmentation. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 2690–2698, 2011.

2010

[560] M. Alamgir, M. Grosse-Wentrup, Y. Altun. Multitask Learning for Brain-Computer Interfaces. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*. Vol. 9. JMLR Workshop and Conference Proceedings, pages 17–24, 2010 (cited on pages 9, 30).

[561] M. Alamgir, U. von Luxburg. Multi-agent random walks for local clustering. In *Proceedings of the IEEE International Conference on Data Mining (ICDM 2010)*, pages 18–27, 2010.

[562] P. Daniusis, D. Janzing, J. Mooij, J. Zscheischler, B. Steudel, K. Zhang, B. Schölkopf. Inferring deterministic causal relations. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pages 143–150, 2010 (cited on pages 6, 18).

[563] A. Erkan, G. Camps-Valls, Y. Altun. Semi-supervised Remote Sensing Image Classification via Maximum Entropy. In *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)*, pages 313–318, 2010.

[564] A. Erkan, O. Kroemer, R. Detry, Y. Altun, J. Piater, J. Peters. Learning Probabilistic Discriminative Models of Grasp Affordances under Limited Supervision. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)*, pages 1586–1591, 2010.

[565] A. Erkan, Y. Altun. Semi-supervised Learning via Generalized Maximum Entropy. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*. Vol. 9. JMLR Workshop and Conference Proceedings, pages 209–216, 2010.

[566] M. Gomez Rodriguez, J. Leskovec, A. Krause. Inferring Networks of Diffusion and Influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010)*, pages 1019–1028, 2010 (cited on page 17).

[567] M. Gomez Rodriguez, J. Peters, J. Hill, B. Schölkopf, A. Gharabaghi, M. Grosse-Wentrup. Closing the sensorimotor loop: Haptic feedback facilitates decoding of arm movement imagery. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC 2010)*, pages 121–126, 2010.

[568] S. Harmeling, S. Sra, M. Hirsch, B. Schölkopf. Multiframe Blind Deconvolution, Super-Resolution, and Saturation Correction via Incremental EM. In *Proceedings of the 17th International Conference on Image Processing (ICIP 2010)*, pages 3313–3316, 2010.

[569] M. Hirsch, B. Schölkopf, M. Habeck. A New Algorithm for Improving the Resolution of Cryo-EM Density Maps. In *Proceedings of the 14th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2010)*. Vol. 6044. Lecture Notes in Bioinformatics, pages 174–188, 2010.

[570] M. Hirsch, S. Sra, B. Schölkopf, S. Harmeling. Efficient Filter Flow for Space-Variant Multiframe Blind Deconvolution. In *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 607–614, 2010 (cited on pages 7, 23).

[571] D. Janzing, P. Hoyer, B. Schölkopf. Telling cause from effect based on high-dimensional observations. In *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, pages 479–486, 2010 (cited on page 18).

[572] D. Kim, S. Sra, I. Dhillon. A scalable trust-region algorithm with application to mixed-norm regression. In *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, pages 519–526, 2010.

[573] J. Kober, K. Mülling, O. Krömer, C. Lampert, B. Schölkopf, J. Peters. Movement Templates for Learning of Hitting and Batting. In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA 2010)*, pages 853–858, 2010.

[574] O. Kroemer, R. Detry, J. Piater, J. Peters. Grasping with Vision Descriptors and Motor Primitives. In *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2010)*, pages 47–54, 2010 (cited on page 8).

[575] O. Kroemer, R. Detry, J. Piater, J. Peters. Adapting Preshaped Grasping Movements Using Vision Descriptors. In *From Animals to Animats 11: Eleventh International Conference on the Simulation of Adaptive Behavior (SAB 2010)*, pages 156–166, 2010 (cited on page 8).

[576] C. Lampert. An efficient divide-and-conquer cascade for nonlinear object detection. In *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 1022–1029, 2010.

[577] C. Lampert, O. Kroemer. Weakly-Paired Maximum Covariance Analysis for Multimodal Dimensionality Reduction and Transfer Learning. In *Proceedings of the 11th European Conference on Computer Vision (ECCV 2010)*, pages 566–579, 2010.

[578] J. Leiva, S. Martens. MLSP Competition, 2010: Description of first place method. In *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)*, pages 112–113, 2010.

[579] K. Mülling, J. Kober, J. Peters. Simulating Human Table Tennis with a Biomimetic Robot Setup. In *From Animals to Animats 11: Eleventh International Conference on the Simulation of Adaptive Behavior (SAB 2010)*, pages 273–282, 2010.

[580] K. Mülling, J. Kober, J. Peters. A biomimetic approach to robot table tennis. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)*, pages 1921–1926, 2010.

[581] K. Mülling, J. Kober, J. Peters. Learning Table Tennis with a Mixture of Motor Primitives. In *Proceedings of the 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2010)*, pages 411–416, 2010 (cited on pages 8, 27).

[582] D. Nguyen-Tuong, J. Peters. Using Model Knowledge for Learning Inverse Dynamics. In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA 2010)*, pages 2677–2682, 2010.

[583] D. Nguyen-Tuong, J. Peters. Incremental Sparsification for Real-time Online Model Learning. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*. Vol. 9. JMLR Workshop and Conference Proceedings, pages 557–564, 2010 (cited on page 27).

[584] H. Nickisch, C. Rasmussen. Gaussian Mixture Modeling with Gaussian Process Latent Variable Models. In *Pattern Recognition: 32nd DAGM Symposium*, pages 271–282, 2010.

[585] J. Peters, D. Janzing, B. Schölkopf. Identifying Cause and Effect on Discrete Data using Additive Noise Models. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*. Vol. 9. JMLR Workshop and Conference Proceedings, pages 597–604, 2010 (cited on page 18).

[586] J. Peters, K. Mülling, Y. Altun. Relative Entropy Policy Search. In *Proceedings of the Twenty-Fourth National Conference on Artificial Intelligence (AAAI 2010)*, pages 1607–1612, 2010 (cited on pages 9, 28).

[587] V. Schwamberger, M. Franz. Simple algorithmic modifications for improving blind steganalysis performance. In *Proceedings of the 12th ACM workshop on Multimedia and Security (MM&Sec 2010)*, pages 225–230, 2010.

[588] V. Schwamberger, P. Le, B. Schölkopf, M. Franz. The Influence of the Image Basis on Modeling and Steganalysis Performance. In *Information Hiding:12th International Conference (IH 2010)*, pages 133–144, 2010.

[589] B. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, G. Lanckriet. Non-parametric estimation of integral probability metrics. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2010)*, pages 1428–1432, 2010.

[590] B. Steudel, D. Janzing, B. Schölkopf. Causal Markov condition for submodular information measures. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 2010)*, pages 464–476, 2010 (cited on page 6).

[591] K. Zhang, A. Hyvärinen. Source Separation and Higher-Order Causal Analysis of MEG and EEG. In *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Sixth Conference (UAI 2010)*, pages 709–716, 2010.

[592] K. Zhang, B. Schölkopf, D. Janzing. Invariant Gaussian Process Latent Variable Models and Application in Causal Discovery. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)*, pages 717–724, 2010.

[593] M.-L. Zhang, K. Zhang. Multi-Label Learning by Exploiting Label Dependency. In *Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2010)*, pages 999–1008, 2010.

[594] M. Alvarez, J. Peters, B. Schölkopf, N. Lawrence. Switched Latent Force Models for Movement Segmentation. In *Advances in neural information processing systems 23 (NIPS 2010)*, pages 55–63, 2010 (cited on page 27).

[595] A. Boularias, B. Chaib-Draa. Bootstrapping Apprenticeship Learning. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pages 289–297, 2010.

[596] S. Chiappa, J. Peters. Movement extraction by detecting dynamics switches and repetitions. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pages 388–396, 2010 (cited on page 27).

[597] S. Harmeling, M. Hirsch, B. Schölkopf. Space-Variant Single-Image Blind Deconvolution for Removing Camera Shake. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pages 829–837, 2010 (cited on pages 7, 23).

[598] J. Mooij, O. Stegle, D. Janzing, K. Zhang, B. Schölkopf. Probabilistic latent variable models for distinguishing between cause and effect. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pages 1687–1695, 2010.

[599] U. von Luxburg, A. Radl, M. Hein. Getting lost in space: Large sample analysis of the resistance distance. In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, pages 2622–2630, 2010 (cited on page 13).

[600] J. Mooij, D. Janzing. Distinguishing between cause and effect. In *Proceedings of the Workshop: Causality: Objectives and Assessment at NIPS 2008*. Vol. 6. JMLR Workshop and Conference Proceedings, pages 147–156, 2010.

[601] I. Guyon, D. Janzing, B. Schölkopf. Causality: Objectives and Assessment. In *Proceedings of the Workshop: Causality: Objectives and Assessment at NIPS 2008*. Vol. 6. JMLR Workshop and Conference Proceedings, pages 1–42, 2010.

[602] L. Song, A. Gretton, C. Guestrin. Nonparametric Tree Graphical Models. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*. Vol. 9. JMLR Workshop and Conference Proceedings, pages 765–772, 2010 (cited on pages 5, 15).

[603] S. Danafar, A. Gretton, J. Schmidhuber. Characteristic Kernels on Structured Domains Excel in Robotics and Human Action Recognition. In *Machine Learning and Knowledge Discovery in Databases, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010)*. Vol. LNCS 6321, pages 264–279, 2010.

[604] C. Widmer, J. Leiva, Y. Altun, G. Rätsch. Leveraging Sequence Classification by Taxonomy-based Multitask Learning. In *Proceedings of the 14th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2010)*. Vol. LNCS 6044, pages 522–534, 2010.

[605] A. Boularias, B. Chaib-Draa. Apprenticeship learning via soft local homomorphisms. In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA 2010)*, pages 2971–2976, 2010.

[606] K. Zhang, A. Hyvärinen. Distinguishing Causes from Effects using Nonlinear Acyclic Causal Models. In *Proceedings of the Workshop: Causality: Objectives and Assessment at NIPS 2008*. Vol. 6. JMLR Workshop and Conference Proceedings, pages 157–164, 2010.

2009

[607] A. Barbero, M. Franz, W. Drongelen, J. Dorronsoro, B. Schölkopf, M. Grosse-Wentrup. Implicit Wiener Series Analysis of Epileptic Seizure Recordings. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)*, pages 5304–5307, 2009.

[608] M. Blaschko, A. Gretton. Learning Taxonomies by Dependence Maximization. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 153–160, 2009.

[609] M. Blaschko, C. Lampert. Object Localization with Global and Local Context Kernels. In *Proceedings of the British Machine Vision Conference 2009 (BMVC 2009)*, pages 1–11, 2009.

[610] S. Chiappa, H. Saigo, K. Tsuda. A Bayesian Approach to Graph Regression with Relevant Subgraph Selection. In *Proceedings of the 2009 SIAM International Conference on Data Mining (SDM 2009)*, pages 295–304, 2009.

[611] S. Chiappa, J. Kober, J. Peters. Using Bayesian Dynamical Systems for Motion Template Libraries. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 297–304, 2009.

[612] D. Corfield. Varieties of Justification in Machine Learning. In *Proceedings of Multiplicity and Unification in Statistics and Probability*, pages 1–10, 2009.

[613] P. Dhillon, S. Nowozin, C. Lampert. Combining appearance and motion for human action classification in videos. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2009 (CVPR Workshops 2009)*, pages 22–29, 2009.

[614] K. Fukumizu, B. Sriperumbudur, A. Gretton, B. Schölkopf. Characteristic Kernels on Groups and Semigroups. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 473–480, 2009.

[615] P. Gehler, S. Nowozin. Let the Kernel Figure it Out: Principled Learning of Pre-processing for Kernel Classifiers. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 2836–2843, 2009.

[616] E. Georgii, K. Tsuda, B. Schölkopf. Multi-way set enumeration in real-valued tensors. In *Proceedings of the 2nd Workshop on Data Mining using Matrices and Tensors (DMMT 2009)*, pages 32–41, 2009.

[617] M. Gomez Rodriguez, J. Kober, B. Schölkopf. Denoising photographs using dark frames optimized by quadratic programming. In *Proceedings of the First IEEE International Conference on Computational Photography (ICCP 2009)*, pages 1–9, 2009.

[618] M. Grosse-Wentrup. Understanding Brain Connectivity Patterns during Motor Imagery for Brain-Computer Interfacing. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 561–568, 2009.

[619] H. Hachiya, J. Peters, M. Sugiyama. Efficient Sample Reuse in EM-Based Policy Search. In *Machine Learning and Knowledge Discovery in Databases: European Conference ECML PKDD 2009*, pages 469–484, 2009.

[620] S. Harmeling, M. Hirsch, S. Sra, B. Schölkopf. Online blind deconvolution for astronomical imaging. In *Proceedings of the First IEEE International Conference Computational Photography (ICCP 2009)*, pages 1–7, 2009 (cited on pages 7, 23).

[621] J. Hill, J. Farquhar, S. Martens, F. Biessmann, B. Schölkopf. Effects of Stimulus Type and of Error-Correcting Code Design on BCI Speller Performance. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 665–672, 2009.

[622] M. Hoffman, N. Freitas, A. Doucet, J. Peters. An Expectation Maximization Algorithm for Continuous Markov Decision Processes with Arbitrary Reward. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*. Vol. 5. JMLR Workshop and Conference Proceedings, pages 232–239, 2009.

[623] M. Hofmann, B. Schölkopf, I. Bezrukov, N. Cahill. Incorporating Prior Knowledge on Class Probabilities into Local Similarity Measures for Intermodality Image Registration. In *Proceedings of the MICCAI 2009 Workshop on Probabilistic Models for Medical Image Analysis (PMMIA 2009)*, pages 220–231, 2009.

[624] P. Hoyer, D. Janzing, J. Mooij, J. Peters, B. Schölkopf. Nonlinear causal discovery with additive noise models. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 689–696, 2009 (cited on pages 6, 18).

[625] D. Janzing, J. Peters, J. Mooij, B. Schölkopf. Identifying confounders using additive noise models. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009)*, pages 249–257, 2009 (cited on pages 6, 18).

[626] S. Jegelka, A. Gretton, B. Schölkopf, B. Sriperumbudur, U. von Luxburg. Generalized Clustering via Kernel Embeddings. In *KI 2009: AI and Automation, Lecture Notes in Computer Science, Vol. 5803*, pages 144–152, 2009.

[627] S. Jegelka, S. Sra, A. Banerjee. Approximation Algorithms for Tensor Clustering. In *Algorithmic Learning Theory: 20th International Conference (ALT 2009)*, pages 368–383, 2009.

[628] H. Kashima, S. Oyama, Y. Yamanishi, K. Tsuda. On Pairwise Kernels: An Efficient Alternative and Generalization Analysis. In *Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference PAKDD 2009*, pages 1030–1037, 2009.

[629] H. Kashima, T. Kato, Y. Yamanishi, M. Sugiyama, K. Tsuda. Link Propagation: A Fast Semi-supervised Learning Algorithm for Link Prediction. In *Proceedings of the 2009 SIAM International Conference on Data Mining (SDM 2009)*, pages 1099–1110, 2009.

[630] J. Kober, J. Peters. Learning motor primitives for robotics. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA 2009)*, pages 2112–2118, 2009.

[631] J. Kober, J. Peters. Policy Search for Motor Primitives in Robotics. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 849–856, 2009.

[632] R. Kondor, N. Shervashidze, K. Borgwardt. The graphlet spectrum. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, pages 529–536, 2009.

[633] O. Kroemer, R. Detry, J. Piater, J. Peters. Active learning using mean shift optimization for robot grasping. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)*, pages 2610–2615, 2009.

[634] B. Kulis, S. Sra, I. Dhillon. Convex Perturbations for Scalable Semidefinite Programming. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*. Vol. 5. JMLR Workshop and Conference Proceedings, pages 296–303, 2009.

[635] C. Lampert. Detecting Objects in Large Image Collections and Videos by Efficient Subimage Retrieval. In *Proceedings of the Twelfth IEEE International Conference on Computer Vision (ICCV 2009)*, pages 987–994, 2009.

[636] C. Lampert, J. Peters. Active Structured Learning for High-Speed Object Detection. In *Pattern Recognition: 31st DAGM Symposium (DAGM 2009)*, pages 221–231, 2009.

[637] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N. Cahill, B. Schölkopf. Learning Similarity Measure for Multi-Modal 3D Image Registration. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 186–193, 2009.

[638] C. Lippert, O. Stegle, Z. Ghahramani, K. Borgwardt. A kernel method for unsupervised structured network inference. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*. Vol. 5. JMLR Workshop and Conference Proceedings, pages 368–375, 2009.

[639] M. Maier, U. von Luxburg, M. Hein. Influence of graph construction on graph-based clustering measures. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1025–1032, 2009.

[640] J. Mooij, B. Kappen. Bounds on marginal probability distributions. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1105–1112, 2009.

[641] J. Mooij, D. Janzing, J. Peters, B. Schölkopf. Regression by dependence minimization and its application to causal inference in additive noise models. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, pages 745–752, 2009.

[642] G. Neumann, J. Peters. Fitted Q-iteration by Advantage Weighted Regression. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1177–1184, 2009.

[643] G. Neumann, W. Maass, J. Peters. Learning Complex Motions by Sequencing Simpler Motion Templates. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, pages 753–760, 2009.

[644] D. Nguyen-Tuong, B. Schölkopf, J. Peters. Sparse online model learning for robot control with support vector regression. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)*, pages 3121–3126, 2009.

[645] D. Nguyen-Tuong, M. Seeger, J. Peters. Local Gaussian Process Regression for Real Time Online Model Learning and Control. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1193–1200, 2009.

[646] H. Nickisch, M. Seeger. Convex variational Bayesian inference for large scale generalized linear models. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, pages 761–768, 2009.

[647] S. Nowozin, C. Lampert. Global Connectivity Potentials for Random Field Models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 818–825, 2009.

[648] S. Nowozin, S. Jegelka. Solution Stability in Linear Programming Relaxations: Graph Partitioning and Unsupervised Learning. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, pages 769–776, 2009.

[649] J. Peters, D. Janzing, A. Gretton, B. Schölkopf. Detecting the Direction of Causal Time Series. In *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, pages 801–808, 2009.

[650] G. Schweikert, C. Widmer, B. Schölkopf, G. Rätsch. An Empirical Analysis of Domain Adaptation Algorithms for Genomic Sequence Analysis. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1433–1440, 2009 (cited on page 10).

[651] M. Seeger, H. Nickisch, R. Pohmann, B. Schölkopf. Bayesian Experimental Design of Magnetic Resonance Imaging Sequences. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1441–1448, 2009 (cited on page 7).

[652] Y. Seldin, N. Tishby. PAC-Bayesian Generalization Bound for Density Estimation with Application to Co-clustering. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*. Vol. 5. JMLR Workshop and Conference Proceedings, pages 472–479, 2009.

[653] H. Shen, K. Hüper. Block Jacobi-type methods for non-orthogonal joint diagonalisation. In *Proceedings of the 34th International Conference on Acoustics, Speech, and Signal Processing (ICASSP09)*, pages 3285–3288, 2009.

[654] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, K. Borgwardt. Efficient Graphlet Kernels for Large Graph Comparison. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*. Vol. 5. JMLR Workshop and Conference Proceedings, pages 488–495, 2009.

[655] S. Sra, D. Kim, I. Dhillon, B. Schölkopf. A new non-monotonic algorithm for PET image reconstruction. In *Proceedings of the IEEE Nuclear Science Symposium Conference Record (NSS/MIC 2009)*. Vol. 2009, pages 2500–2502, 2009.

[656] O. Stegle, K. Denby, D. Wild, Z. Ghahramani, K. Borgwardt. A robust Bayesian two-sample test for detecting intervals of differential gene expression in microarray time series. In *Proceedings of the 13th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2009)*, pages 201–216, 2009.

[657] O. Stegle, K. Denby, D. Wild, S. McHattie, A. Mead, Z. Ghahramani, K. Borgwardt. Discovering Temporal Patterns of Differential Gene Expression in Microarray Time Series. In *Proceedings of the German Conference on Bioinformatics 2009 (GCB 2009)*, pages 133–142, 2009.

[658] F. Steinke, M. Hein. Non-parametric Regression between Riemannian Manifolds. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1561–1568, 2009.

[659] M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A. Smola, L. Song, P. Yu, X. Yan, K. Borgwardt. Near-optimal supervised feature selection among frequent subgraphs. In *Proccedings of the 2009 SIAM Conference on Data Mining (SDM 2009)*, pages 1076–1087, 2009.

[660] C. Walder, B. Schölkopf. Diffeomorphic Dimensionality Reduction. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1713–1720, 2009.

[661] X. Zhang, L. Song, A. Gretton, A. Smola. Kernel Measures of Independence for Non-IID Data. In *Advances in neural information processing systems 21 (NIPS 2008)*, pages 1937–1944, 2009.

[662] M. Blaschko, J. Shelton, A. Bartels. Augmenting Feature-driven fMRI Analyses: Semi-supervised learning and resting state activity. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 126–134, 2009.

[663] L. Cayton. Efficient Bregman Range Search. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 243–251, 2009.

[664] A. Gretton, K. Fukumizu, Z. Harchaoui, B. Sriperumbudur. A Fast, Consistent Kernel Two-Sample Test. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 673–681, 2009.

[665] N. Shervashidze, K. Borgwardt. Fast subtree kernels on graphs. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 1660–1668, 2009.

[666] B. Sriperumbudur, K. Fukumizu, A. Gretton, G. Lanckriet, B. Schölkopf. Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 1750–1758, 2009.

[667] R. Tillman, A. Gretton, P. Spirtes. Nonlinear directed acyclic structure learning with weakly additive noise models. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, pages 1847–1855, 2009.

[668] S. Jegelka, J. Bilmes. Notes on Graph Cuts with Submodular Edge Weights. In *NIPS 2009 Workshop on Discrete Optimization in Machine Learning: Submodularity, Sparsity & Polyhedra (DISCML)*, pages 1–6, 2009.

[669] S. Martens, J. Farquhar, J. Hill, B. Schölkopf. Graphical models for decoding in BCI visual speller systems. In *4th International IEEE EMBS Conference on Neural Engineering (NER 2009)*, pages 470–473, 2009.

[670] C. Walder, M. Breidt, H. Bülthoff, B. Schölkopf, C. Curio. Markerless 3D Face Tracking (DAGM 2009). In *Pattern Recognition, Lecture Notes in Computer Science, Vol. 5748*, pages 41–50, 2009.

1.6.5 Book Chapters

2015

[671] D. Janzing, B. Steudel, N. Shajarisales, B. Schölkopf. Justifying Information-Geometric Causal Inference. In. *Measures of Complexity: Festschrift for Alexey Chervonenkis*. Ed. by V. Vovk, H. Papadopoulos, A. Gammerman. Springer. Chap. 18, pages 253–265, 2015.

[672] G. Charpiat, M. Hofmann, B. Schölkopf. Kernel methods in medical imaging. In. *Handbook of Biomedical Imaging*. Ed. by N. Paragios, J. Duncan, N. Ayache. Springer, Berlin, Germany. Chap. 4, pages 63–81, 2015.

2014

[673] S. Sra. Nonconvex Proximal Splitting with Computational Errors. In. *Regularization, Optimization, Kernels, and Support Vector Machines*. Ed. by J. A. K. Suykens, M. Signoretto, A. Argyriou. CRC Press. Chap. 4, pages 83–102, 2014 (cited on page 16).

[674] T. Schultz, A. Vilanova, R. Brecheisen, G. Kindlmann. Fuzzy Fibers: Uncertainty in dMRI Tractography. In. *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Ed. by C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, H. Hagen. Mathematics + Visualization. Springer. Chap. 8, pages 79–92, 2014.

[675] T. Schultz, A. Fuster, A. Ghosh, R. Deriche, L. Florack, L.-H. Lim. Higher-Order Tensors in Diffusion Imaging. In. *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*. Ed. by C.-F. Westin, A. Vilanova, B. Burgeth. Mathematics + Visualization. Springer, pages 129–161, 2014.

[676] K. Zhang, B. Schölkopf, K. Muandet, Z. Wang, Z. Zhou, C. Persello. Single-Source Domain Adaptation with Target and Conditional Shift. In. *Regularization, Optimization, Kernels, and Support Vector Machines*. Ed. by J. Suykens, M. Signoretto, A. Argyriou. Chapman & Hall/CRC Machine Learning & Pattern Recognition. Chapman and Hall/CRC, Boca Raton, USA. Chap. 19, pages 427–456, 2014 (cited on page 14).

2013

[677] Y. Seldin, B. Schölkopf. On the Relations and Differences between Popper Dimension, Exclusion Dimension and VC-Dimension. In. *Empirical Inference - Festschrift in Honor of Vladimir N. Vapnik*. Ed. by B. Schölkopf, Z. Luo, V. Vovk. Springer. Chap. 6, pages 53–57, 2013 (cited on page 12).

[678] S. Sra. Tractable large-scale optimization in machine learning. In. *Tractability: Practical Approaches to Hard Problems*. Ed. by L. Bordeaux, Y. Hamadi, P. Kohli, R. Mateescu. Cambridge University Press. Chap. 7, pages 202–230, 2013 (cited on page 16).

[679] M. Grosse-Wentrup, B. Schölkopf. A Review of Performance Variations in SMR-Based Brain–Computer Interfaces (BCIs). In. *Brain-Computer Interface Research*. Ed. by C. Guger, B. Allison, G. Edlinger. SpringerBriefs in Electrical and Computer Engineering. Springer. Chap. 4, pages 39–51, 2013.

[680] B. Schölkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, J. Mooij. Semi-supervised learning in causal and anticausal settings. In. *Empirical Inference - Festschrift in Honor of Vladimir N. Vapnik*. Ed. by B. Schölkopf, Z. Luo, V. Vovk. Springer. Chap. 13, pages 129–141, 2013.

2012

[681] M. Toussaint, A. Storkey, S. Harmeling. Expectation-Maximization methods for solving (PO)MDPs and optimal control problems. In. *Inference and Learning in Dynamic Models*. Ed. by D. Barber, A. Cemgil, S. Chiappa. Cambridge University Press, Cambridge, UK, 2012.

[682] O. Sigaud, J. Peters. Robot Learning. In. *Encyclopedia of the Sciences of Learning*. Ed. by N. Seel. Springer, Berlin, Germany, 2012.

[683] J. Kober, J. Peters. Reinforcement Learning in Robotics: A Survey. In. *Reinforcement Learning*. Ed. by M. Wiering, M. Otterlo. Vol. 12. Springer, Berlin, Germany, pages 579–610, 2012 (cited on page 8).

[684] T. Schultz, A. Fuster, A. Ghosh, R. Deriche, L. Florack, L. Lim. Higher-Order Tensors in Diffusion MRI. In. *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*. Ed. by C. Westin, A. Vilanova, B. Burgeth. Springer, 2012.

[685] M. Habeck. Inferential structure determination from NMR data. In. *Bayesian methods in structural bioinformatics*. Ed. by T. Hamelryck, K. Mardia, J. Ferkinghoff-Borg. Springer, New York, pages 287–312, 2012.

2011

[686] J. Peters, R. Tedrake, N. Roy, J. Morimoto. Robot Learning. In. *Encyclopedia of Machine Learning*. Ed. by C. Sammut, G. I. Webb. Encyclopedia of machine learning. Springer, New York, NY, USA, pages 865–869, 2011.

[687] M. Schmidt, D. Kim, S. Sra. Projected Newton-type methods in machine learning. In. *Optimization for Machine Learning*. Ed. by S. Sra, S. Nowozin, S. J. Wright. MIT Press, Cambridge, MA, USA, pages 305–330, 2011.

[688] U. von Luxburg, B. Schölkopf. Statistical Learning Theory: Models, Concepts, and Results. In. *Handbook of the History of Logic, Vol. 10: Inductive Logic*. Ed. by D. Gabbay, S. Hartmann, J. Woods. Vol. 10. Elsevier North Holland, Amsterdam, Netherlands, pages 651–706, 2011.

[689] K. Borgwardt. Kernel Methods in Bioinformatics. In. *Handbook of Statistical Bioinformatics*. Ed. by H.-S. Lu, B. Schölkopf, H. Zhao. Springer Handbooks of Computational Statistics ; 3. Springer, Berlin, Germany, pages 317–334, 2011.

[690] P. Rosas, F. Wichmann. Cue Combination: Beyond Optimality. In. *Sensory Cue Integration*. Ed. by J. Trommershäuser, K. Körding, M. Landy. Oxford University Press, pages 144–152, 2011.

[691] K. Ihme, T. Zander. What You Expect Is What You Get? Potential Use of Contingent Negative Variation for Passive BCI Systems in Gaze-Based HCI. In. *Affective Computing and Intelligent Interaction*. Ed. by S. D'Mello, A. Graesser, B. Schuller, J.-C. Martin. Vol. 6975. Lecture Notes in Computer Science. Springer, Berlin, Germany, pages 447–456, 2011.

2010

[692] G. Charpiat, I. Bezrukov, M. Hofmann, Y. Altun, B. Schölkopf. Machine Learning Methods for Automatic Image Colorization. In. *Computational Photography: Methods and Applications*. Ed. by R. Lukac. Digital Imaging and Computer Vision. CRC Press, Boca Raton, FL, USA, pages 395–418, 2010.

[693] R. Detry, E. Baseski, M. Popovic, Y. Touati, N. Krüger, O. Kroemer, J. Peters, J. Piater. Learning Continuous Grasp Affordances by Sensorimotor Exploration. In. *From Motor Learning to Interaction Learning in Robots*. Ed. by J. P. Sigaud O. Vol. 264. Studies in Computational Intelligence. Springer, Berlin, Germany, pages 451–465, 2010.

[694] J. Kober, B. Mohler, J. Peters. Imitation and Reinforcement Learning for Motor Primitives with Perceptual Coupling. In. *From Motor Learning to Interaction Learning in Robots*. Ed. by J. P. Sigaud O. Vol. 264. Studies in Computational Intelligence. Springer, Berlin, Germany, pages 209–225, 2010.

[695] D. Nguyen-Tuong, M. Seeger, J. Peters. Real-Time Local GP Model Learning. In. *From Motor Learning to Interaction Learning in Robots*. Ed. by O. Sigaud, J. Peters. Vol. 264. Studies in Computational Intelligence. Springer, Berlin, Germany, pages 193–207, 2010.

[696] J. Peters, J. Bagnell. Policy Gradient Methods. In. *Encyclopedia of Machine Learning*. Ed. by G. I. W. Sammut C. Springer, Berlin, Germany, pages 774–776, 2010.

[697] C. Walder, M. Breidt, H. Bülthoff, B. Schölkopf, C. Curio. Markerless tracking of Dynamic 3D Scans of Faces. In. *Dynamic Faces: Insights from Experiments and Computation*. Ed. by C. Curio, H. Bülthoff, M. Giese. MIT Press, Cambridge, MA, USA, pages 255–276, 2010.

2009

[698] Y. Altun. Large Margin Methods for Part of Speech Tagging. In. *Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods*. Ed. by J. Keshet, S. Bengio. Wiley, Hoboken, NJ, USA, pages 141–160, 2009.

[699] A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, B. Schölkopf. Covariate shift and local learning by distribution matching. In. *Dataset Shift in Machine Learning*. Ed. by J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, N. Lawrence. MIT Press, Cambridge, MA, USA, pages 131–160, 2009.

[700] S. Sra, A. Banerjee, J. Ghosh, I. Dhillon. Text Clustering with Mixture of von Mises-Fisher Distributions. In. *Text mining: classification, clustering, and applications*. Ed. by A. Srivastava, M. Sahami. Chapman & Hall/CRC data mining and knowledge discovery series. CRC Press, Boca Raton, FL, USA, pages 121–161, 2009.

[701] K. Tsuda. Data Mining for Biologists. In. *Biological Data Mining in Protein Interaction Networks*. Ed. by X. Li, S.-K. Ng. Medical Information Science Reference, Hershey, PA, USA, pages 14–27, 2009.

1.6.6 Theses

PhD Theses

[702] R. Köhler. Advances in computational imaging: Benchmarking Deblurring Algorithms, Deep Neural Inpainting, Depth Estimation from Light Fields. PhD thesis. University of Tübingen, Germany, 2016.

[703] D. Grimm. easyGWAS: An Integrated Computational Framework for Advanced Genome-Wide Association Studies. PhD thesis. University of Tübingen, 2015.

[704] E. Sgouritsa. Causal Discovery Beyond Conditional Independences. PhD thesis. University of Tübingen, Germany, 2015.

[705] C. Schuler. Machine Learning Approaches to Image Deconvolution. PhD thesis. University of Tübingen, Germany, 2015.

[706] K. Muandet. From Points to Probability Measures: A Statistical Learning on Distributions with Kernel Mean Embedding. PhD thesis. University of Tübingen, Germany, 2015.

[707] A. Loktyushin. Blind Retrospective Motion Correction of MR Images. PhD thesis. University of Tübingen, Germany, 2015.

[708] B. Rakitsch. Modeling the polygenic architecture of complex traits. PhD thesis. University of Tübingen, 2014.

[709] J. Zscheischler. A global analysis of extreme events and consequences for the terrestrial carbon cycle. PhD thesis. ETH Zurich, Switzerland, 2014.

[710] M. Alamgir. Analysis of Distance Functions in Graphs. PhD thesis. University of Hamburg, Germany, 2014.

[711] C. Lippert. Linear mixed models for genome-wide association studies. PhD thesis. University of Tübingen, Germany, 2013.

[712] Z. Wang. Intention Inference and Decision Making with Hierarchical Gaussian Process Dynamics Models. PhD thesis. Technical University Darmstadt, Germany, 2013.

[713] K. Mülling. Modeling and Learning Complex Motor Tasks: A case study on Robot Table Tennis. PhD thesis. Technical University Darmstadt, Germany, 2013.

[714] H. Burger. Modelling and Learning Approaches to Image Denoising. PhD thesis. University of Tübingen, Germany, 2013.

[715] J. Peters. Restricted structural equation models for causal inference. PhD thesis. ETH Zurich, Switzerland, 2012.

[716] J. Kober. Learning Motor Skills: From Algorithms to Robot Experiments. PhD thesis. Technische Universität Darmstadt, Germany, 2012.

[717] N. Shervashidze. Scalable graph kernels. PhD thesis. University of Tübingen, Germany, 2012.

[718] M. Hirsch. Blind Deconvolution in Scientific Imaging & Computational Photography. PhD thesis. University of Tübingen, Germany, 2012 (cited on page 7).

[719] S. Jegelka. Combinatorial Problems with Submodular Coupling in Machine Learning and Computer Vision. PhD thesis. ETH Zürich, Switzerland, 2012.

[720] M. Gomez Rodriguez. Structure and Dynamics of Diffusion Networks. PhD thesis. Department of Electrical Engineering, Stanford University, 2012.

[721] D. Nguyen-Tuong. Model Learning in Robot Control. PhD thesis. Albert-Ludwigs-Universität Freiburg, Germany, 2011.

[722] H. Nickisch. Bayesian Inference and Experimental Design for Large Generalised Linear Models. PhD thesis. Technische Universität Berlin, Berlin, Germany, 2010.

[723] M. Hofmann. Quantitative Positron Emission Tomography with a Combined PET/MR System. PhD thesis. University of Oxford, UK, 2010.

[724] J. Macke. Population Coding in the Visual System: Statistical Methods and Theory. PhD thesis. University of Tübingen, Germany, 2010.

[725] M. Maier. Clustering with Neighborhood Graphs. PhD thesis. Universität des Saarlandes, Saarbrücken, Germany, 2010.

[726] S. Martens. A wider view on encoding and decoding in the visual brain-computer interface speller system. PhD thesis. University of Tübingen, Germany, 2010.

[727] G. Schweikert. Accurate Prediction of Protein-Coding Genes with Discriminative Learning Techniques. PhD thesis. Technische Universität Berlin, Germany, 2010.

[728] E. Georgii. Structural and Relational Data Mining for Systems Biology Applications. PhD thesis. University of Tübingen, Germany, 2010.

[729] S. Gerwinn. Bayesian Methods for Neural Data Analysis. PhD thesis. University of Tübingen, Germany, 2010.

[730] M. Blaschko. Kernel Methods in Computer Vision:Object Localization, Clustering,and Taxonomy Discovery. PhD thesis. Technische Universität Berlin, Berlin, Germany, 2009.

[731] P. Gehler. Kernel Learning Approaches for Image Classification. PhD thesis. Universität des Saarlandes, Saarbrücken, Germany, 2009.

[732] S. Nowozin. Learning with Structured Data: Applications to Computer Vision. PhD thesis. Technische Universität Berlin, Germany, 2009.

[733] F. Steinke. From Differential Equations to Differential Geometry: Aspects of Regularisation in Machine Learning. PhD thesis. Universität des Saarlandes, Saarbrücken, Germany, 2009.

Master Theses

[734] A. Ibarra Chaoul. The effect of frowning on attention. MA thesis. Graduate Training Centre of Neuroscience, University of Tübingen, Germany, 2015.

[735] M. Hohmann. A Cognitive Brain-Computer Interface for Patients with Amyotrophic Lateral Sclerosis. MA thesis. Graduate Training Centre of Neuroscience, University of Tübingen, Germany, 2015.

[736] L. Casarsa de Azevedo. Causal Inference in Neuroimaging. MA thesis. Graduate Training Centre of Neuroscience, University of Tübingen, Germany, 2015.

[737] M. Gao. Sequential Image Deconvolution Using Probabilistic Linear Algebra. MA thesis. Technical University of Munich, Germany, 2015.

[738] E. Lacosse. The Feasibility of Causal Discovery in Complex Systems: An Examination of Climate Change Attribution and Detection. MA thesis. Graduate Training Centre of Neuroscience, University of Tübingen, Germany, 2014.

[739] B. Huang. Causal Discovery in the Presence of Time-Dependent Relations or Small Sample Size. MA thesis. Graduate Training Centre of Neuroscience, University of Tübingen, Germany, 2014.

[740] N. Shajarisales. A Novel Causal Inference Method for Time Series. MA thesis. University of Tübingen, Germany, 2014.

[741] A. Lang. Crowdsourcing for optimisation of deconvolution methods via an iPhone application. MA thesis. Hochschule Reutlingen, Germany, 2011.

Diploma Theses

[742] N. Schmeißer. Development of advanced methods for improving astronomical images. University of Tübingen, Germany, 2014.

[743] M. Schober. Camera-specific Image Denoising. University of Tübingen, Germany, 2013.

[744] J. Hooge. Automatische Seitenkettenzuordnung zur NMR Proteinstrukturaufklärung mittels ganzzahliger linearer Programmierung. University of Tübingen, Germany, 2012.

[745] E. Klenske. Nonparametric System Identification and Control for Periodic Error Correction in Telescopes. Universität Stuttgart, Germany, 2012.

[746] F. Mantlik. Quantitative Evaluation of MR-based Attenuation Correction for Positron Emission Tomography (PET). Universität Mannheim, Germany, 2010.

[747] J. Shelton. Semi-supervised Subspace Learning and Application to Human Functional Magnetic Brain Resonance Imaging Data. University of Tübingen, Germany, 2010.

[748] J. Zscheischler. Inferring High-Dimensional Causal Relations using Free Probability Theory. Humboldt Universität Berlin, Germany, 2010.

[749] B. Rakitsch. Finding Gene-Gene Interactions using Support Vector Machines. University of Tübingen, Germany, 2010.

[750] R. Köhler. Detecting the mincut in sparse random graphs. University of Tübingen, Germany, 2010.

[751] K. Mülling. Motor Control and Learning in Table Tennis. University of Tübingen, Gerrmany, 2009.

[752] P. Drewe. Hierarchical Clustering and Density Estimation Based on k-nearest-neighbor graphs. University of Tübingen, Germany, 2009.