Review Article · DOI: 10.2478/s13230-012-0022-3 · JBR · 3(2) · 2012 · 75-91

From Humans to Humanoids: the Optimal Control Framework

Serena Ivaldi[1,*], Olivier Sigaud[1,†], Bastien Berret[2,‡], Francesco Nori[2,§]

1 Université Pierre et Marie Curie, Institut des Systèmes Intelligents et de Robotique - CNRS UMR 7222, Pyramide Tour 55 - Boîte Courrier 173, 4 Place Jussieu, 75252 Paris CEDEX 05, France

2 Robotics, Brain and Cognitive Sciences Dept. Istituto Italiano di Tecnologia, Via Morego 30, 16136 Genova, Italy

Received 2012/03/13

Accepted 2012/05/09

Abstract

In the last years of research in cognitive control, neuroscience and humanoid robotics have converged to different frameworks which aim, on one side, at modeling and analyzing human motion, and, on the other side, at enhancing motor abilities of humanoids. In this paper we try to cover the gap between the two areas, giving an overview of the literature in the two fields which concerns the production of movements. First, we survey computational motor control models based on optimality principles; then, we review available implementations and techniques to transfer these principles to humanoid robots, with a focus on the limitations and possible improvements of the current implementations. Moreover, we propose Stochastic Optimal Control as a framework to take into account delays and noise, thus catching the unpredictability aspects typical of both humans and humanoids systems. Optimal Control in general can also easily be integrated with Machine Learning frameworks, thus resulting in a computational implementation of human motor learning. This survey is mainly addressed to roboticists attempting to implement human-inspired controllers on robots, but can also be of interest for researchers in other fields, such as computational motor control.

Keywords

humanoids · human motor control · optimality · stochastic optimal control

1. Introduction

Modern humanoid platforms must be capable of performing complex tasks controlling their entire body and its contacts. Diverse approaches addressing whole-body motion in this context can be found, ranging from hierarchical to model predictive and reactive control, or offline graph-based exploration of states [55, 105, 107, 125, 138, 178]. In addition to safety requirements and compliance during physical interactions [39, 59], robots are also required to perform human-like movements in terms of trajectories, accuracy and reaction to external perturbations, so that the human can have intuitive expectations about the robot's behavior. These constraints generate additional challenges that classical automatic and robotic control are not able to address completely.

Only recently, these issues have been tackled by a new line of research. In this paper, we call it the "**humans to humanoids**" (*H2H*) approach: it consists in studying the Human Motor Control (*HMC*) system in the search for principles of motor control that can be "transferred to", i.e. implemented on, robots. The idea behind the materialization of motor principles is that, by implementing them on humanoid platforms, it is possible to obtain behaviors which outperform traditional classical controls [1, 88], simultaneously providing an experimental verification of the proposed models [6]. While classical control theory has been mainly focused on minimizing tracking errors, rejecting disturbances, minimizing the motion duration, guaranteeing stability, etc., the study of *HMC* may unveil other criteria as the fundamentals for achieving peculiar performances and characteristics which make our motor control so efficient.

This new approach is not driven nor limited by the necessity of robots to exhibit human-like shapes as pure ***replica***. A common approach to reproduce human-like movements consists in recording movements from humans with a motion capture system, then implementing on the robots the necessary software so that it can "replay" similar movements. This research approach faces problems such as the ***correspondence problem*** resulting from the differences in the mechanical structure of the human body and the robotic platform, or the problem of generalizing a whole set of behaviors out of a few recorded trajectories [5, 20, 116]. Though replicating human behaviors and making robots appear more natural in a human environment [137], this approach only works for limited applications, and does not endow the robot with the capability to act and react appropriately in other contexts.

In contrast, in order to make a humanoid more "human", the *H2H* approach aims at endowing the robot with human-inspired motor control principles combined with autonomy and adaptive capabilities [16, 25, 53, 63]. The realization of the fundamental principles of *HMC* is more difficult, but has a greater potential. By addressing the generation of motions from an abstract representation of the goal and through a cost function, the kinematic and dynamic properties of movements may be reproduced without the need for a demonstration.

Several difficulties are usually encountered when attempting *H2H* transfers: in terms of physical and structural impairments, in the technology for actuation and sensing, in the theoretical tools which can be exploited to generate controls and in computational costs. In the past, these constraints restricted the number of applications of such meth-

*E-mail: serena.ivaldi@isir.upmc.fr
†E-mail: olivier.sigaud@isir.upmc.fr
‡E-mail: bastien.berret@iit.it
§E-mail: francesco.nori@iit.it

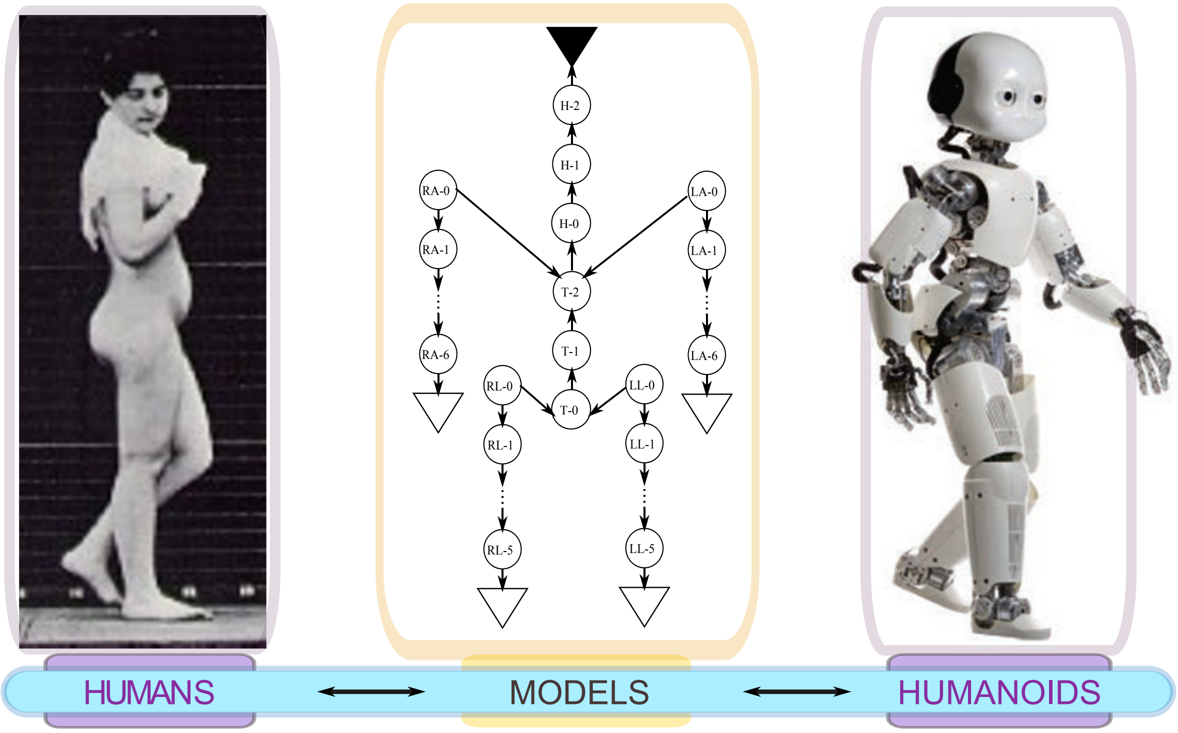

Figure 1. *The paradigm "humans to humanoids" for the transfer of motor principles. Kinematic and dynamic models can be used to describe both human and robotic systems, while optimal control can be used to study and simulate/reproduce their behavior. In the middle, a kinematic model of a whole-body system based on the Enhanced Oriented Graphs [70].*

ods in humanoid robots, but thanks to new technological advances in actuators, sensors [35], and in computational power, *H2H* transfers of optimality principles are gradually emerging. We believe that the main limitation of *H2H* is the lack of a unified general framework which can help roboticists interpreting the neuroscientific results in a way that can be easily implemented on robotic platforms. This paper helps to fill this gap providing an overview of the literature in between and a paradigm to support the "transfer" of these principles towards humanoids. We first survey the state of the art in computational motor control, giving an overview of optimality and adaptation principles in *HMC* and then focus on the transfer of these principles towards humanoids.

The paper is organized as follows. In Section 2, we explain why we endorse Optimal Control (*OC*) as the general framework for implementing such transfer of motor principles. In Section 3, we glance over the organizational and optimality principles of the *HMC* that derive from the constraints on the perception and actuation apparatus of humans, and focus on some computational models giving account of goal-directed movements. We also survey the adaptation mechanisms that make human behavior so robust and efficient in front of unforeseen perturbations, and discuss how the *OC* framework can naturally integrate both control and learning. Then in Section 4 we present the state-of-the-art implementations of *OC* for control and adaptation in robotics and particularly in humanoids. We conclude this literature survey by identifying the forthcoming challenges and giving some insights of future developments. Lastly, in Appendix A we list the acronyms used in this paper and provide the reader a set of references to *OC* in Appendix B.

2. Why (Stochastic) Optimal Control?

There exist several theoretical frameworks to investigate *HMC*. In this section, we explain why we choose *OC* and its stochastic extension as the most appropriate framework to transfer *HMC* principles to the control of humanoids. The interested reader should refer to Appendix B for a definition of an *OC* problem and its stochastic version.

2.1. Frameworks for *HMC*

Overall, three main theoretical frameworks for describing *HMC* have been proposed [56]. In the equilibrium point theory [47], goal-directed movements can be seen as continuous transitions between postures along an equilibrium trajectory. The generated movements themselves result from an imbalance between the spring-like forces generated by a shift in the origin of muscles. In the dynamical system approach to motor control, behavior emerges from regularities of nonlinear dynamical systems [133]. Goal-directed movements are obtained from tuning the parameters of these dynamical systems. Finally, in the internal model framework, goal-directed movements are obtained from an *OC* process using internal models of the system.
These theoretical frameworks have their own merit in catching diverse aspects of *HMC*, and they are likely to give a complementary perspective on the motor control system. However, in this paper we focus only on the latter. Our choice is motivated by a "transferability" criterion: as it will be explained in the following sections, *OC* is the most suited framework to transfer those principles in robotics applications. Thus, in

the rest of the paper, we endorse the view that, in movement production, the Central Nervous System (*CNS*) is optimizing something that remains to be determined.

2.2. Is our behavior "optimal"?

Claiming that our behavior is optimizing something is not equivalent to claiming that it is optimal at any moment, neither that it globally is. Instead, we suggest that optimization can be used to model human movements. Though our sensorimotor system is the product of millions of years of evolution, there are diverse counter-examples in biology where the optimal solution (or at least the one which we believe should be, since optimality is always defined with respect to a certain cost function) is not "attained" by evolution, both from a bio-mechanical point of view and a control point of view, while suboptimal solutions persist [112]. Humans are subject to continuous processes such as learning, adaptation and training, which improve our behavioral performance in terms of stability, accuracy and efficiency [6, 50, 141].

This does not mean that optimality is reached through experience or ontogenesis, as other factors may contribute in movement shaping. Social conventions and habits, for example, play a role in shaping our movements. Sometimes, they cause humans to be stuck in suboptimal behaviors. Roughly speaking, humans can be thought of as optimizing (as opposed to optimal) agents. Therefore, their movements might still be explained by suboptimal solutions. The straddle technique for jumping is a good example of suboptimal movement in humans [38].[1] This example does not contradict the idea that humans are optimizing their movements, but only proves that this optimization might get stuck in local minima.

Moreover, the computational cost of optimization processes may be counter-balanced by some perpetual processes that allow the *CNS* storing the best response in some situation and retrieving it at a much lower cost in similar situations. Recent experiments [53, 80] suggest that the *CNS* could not have just one planning process but rather a set of optimizing planners which can be switched on and off whenever particular changes in the external environment occur, which do not induce an immediate adaptation effect. When the context is evolving, the automatized response may not be optimal anymore, and adaptation mechanisms are evoked.

Overall, the existence of a unique optimizing controller in the *CNS* is still matter of debate. Intuitively, it is more plausible to say that humans are likely to act on the basis of one or more optimality principles, but may choose not "the" optimal solution at all time. These elements suggest that optimality should not be seen as a general property resulting from evolution or learning, but rather as a principled framework to understand the control and adaptation mechanisms.

2.3. The stochastic element

OC has become of great interest in the neuroscientific community [154, 158]. This elegant framework provides a rich set of mathematical tools for modeling computational motor control and explaining empirical human movement data. Among many potential *OC* formalisms, Stochastic Optimal Control (*SOC*) is probably the most appropriate one to study the production of movements in both humans and humanoids. Indeed, from the *HMC* side, diverse stochastic and deterministic *OC* models have been proposed. Remarkably, different models are apparently equally effective in explaining movements in the same or a similar context (see Section 3.2), and it is difficult to assess which is the most representative of human movements, and why. In front of such diversity, *SOC* is general enough to encompass several optimization-based *HMC* theories. The stochastic element indeed is able to address some features such as the variability of motion, the intrinsic noise in the actuation system [155], the ability of the *CNS* to cope with unpredictable events, changing and unknown environments, and delays in the signal transmission. These features can be addressed with difficulty (or not at all) by deterministic models [58]. From an automatic control point of view, deterministic models are rarely useful for real physical systems. Robots, for examples, are complex machines subjects to different sources of noise. Even when acting in highly structured environments, such as an industrial setup, motor control necessarily entails noise filtering, delay compensation and disturbance rejection, and only stochastic models can be used to that scope. Our effort being centered in bridging the gap between humans and humanoids, the ability to deal with the intrinsic and extrinsic noise acting on the systems becomes mandatory. Any cognitive agent creates a model of its own body and the surrounding environment. Though this model can be very precise, it will be nevertheless affected by (unpredictable) errors, which should be properly considered and compensated when planning movements. This unpredictability can be mathematically represented by stochastic systems, i.e. dynamic systems affected by unpredictable noise. Accordingly, optimal planning should be tackled with *SOC*, which computes the best control policy given the effects of unpredictability on the cost function (see Appendix B).

The mathematical tools to tackle *SOC* are usually more complex and generally the solution of such control problems is easy only if certain conditions are met. As it will be shown in 4.5, machine learning techniques can help solving such problems and grant their application in real scenarios.

Hereinafter, we will mostly refer to *OC*, because it is the general framework, and use *SOC* only when the stochastic element is important.

2.4. A criterion of "transferability"

The goal of an *OC* problem is to provide the control signals that will make a system or a process attain a goal while satisfying its constraints and maximizing some performance criterion (or minimizing a cost) which describes its evolution. The formulation of an *OC* control problem (see Appendix B) usually requires a model, i.e. a mathematical description of the system, a list of physical constraints and the specification of the performance/cost function describing the task. These three elements and the mathematical formalism are the core of the *H2H* approach. Indeed, both the human and the humanoid bodies can be modeled by a set of kinematic or dynamic system equations, while their interaction with the environment and their actuation define their set of constraints. Their actions can be described by a set of goals, the latter expressed either as the maximization of some merit function (e.g. a reward), or as the minimization of some cost function (e.g. task error, energy expenditure, etc.).

Neuroscience and robotics benefit from mutual achievements [129], because many problems faced by the primate brain in the control of movement have parallels in robotic motor control, while models and algorithms from automatic control and robotics research can bring useful inspiration, baseline performance, and sometimes direct analogs for neuroscience. Thus, the *OC* framework is naturally suited to implement the computational motor control models provided by neuroscience into

[1] A clarifying example is the introduction of the Fosbury flop. Athletes have been performing High Jumps for years with different styles, the most common being the straddle technique. When the Fosbury flop was introduced, many athletes did not easily switched to the new technique, though it was easier to learn and clearly more advantageous from a biomechanical point of view [36, 37].

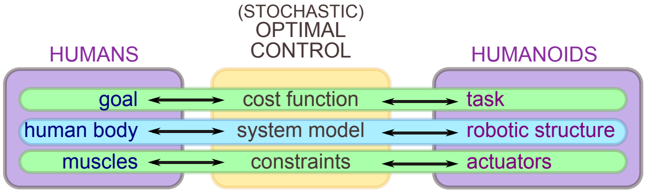

*Figure 2. The framework of optimal control is sufficiently general and abstract to be used to **transfer** control strategies **from humans to humanoids**. Both systems are similar in a sense, since for both we can provide a mathematical description of the plant and its constraints. Once the motor control model along with its cost function is known, it is easy (to a certain extent) to use optimal control theory to transfer motor principles from humans to humanoids, as well as to infer motion criteria from human experimental data using modeling tools which are close to the robotics community.*

controllers for humanoid robots thanks to the common language used to describe both systems (see Figure 2). Additionally, the *OC* formalism is naturally integrated in classical automatic control schemes, which are widely used to model aspects of *HMC* such as feedback and feedforward commands, noise and delays.

Furthermore, *OC* is also adequate for giving a formal account of adaptation capabilities, given the possibility to integrate in the framework adaptive models and adaptive optimization methods [17]. Indeed, *OC* combined with Machine Learning (*ML*) can provide a normative framework for modeling the exceptional dexterity and rapid adaption to changes which characterize *HMC*. Within certain limitations, the framework is also suited for incremental learning, i.e. it can be combined with a developmental approach to humanoid robotics: the system description through models, the cost functions, the controllers, can be all seen as evolving terms in a life-long learning scenario.

3. Optimality principles in human motor control

Human movements show several prominent features [122, 167]. For instance, multi-joint arm trajectories for discrete point-to-point planar movements have roughly straight hand paths, bell-shaped velocity profiles and smooth acceleration [49, 101, 136]. Intuitively, ontogenesis should yield a certain variability in motion patterns. Yet, despite the intrinsic noise of the motor system and the kinematic and muscular redundancy in front of most tasks, observed motion trajectories in many tasks are stereotyped.[2] Why do humans have such invariants? Explaining why the *HMC* system selects a particular movement among such infinite-dimensional possibilities is known as Bernstein's problem [13]. The solution we endorse in this paper consists in claiming that these invariants are the result of an optimization process. Indeed, many *HMC* researchers consider that, in order to perform accurate movements at all times, the *CNS* must be optimizing one or more criteria when deciding how to perform a task and execute a limb trajectory. However, the cues above do not provide a unique answer to the questions regarding the criteria and mechanisms which are at the basis of our movement. Hereinafter, we discuss the control mechanisms of *HMC*, then we review potential optimization criteria.

3.1. Feedback and feedforward

By contrast with industrial and most humanoids robots, the human biomechanical system is light, its sensori-motor apparatus noisy and delayed. Despite these limitations, it can accomplish very accurately complex high-level tasks in presence of disturbances and unpredictable changes in the environment. Accuracy in this case is not based on high stiffness and servo-control, but rather on anticipation and capability to adapt to perturbations, i.e. on a combination of feedback and feedforward control.

Feedforward or open-loop control consists in applying a sequence of controls without monitoring the state of the plant during this sequence (either because the plant is perfectly known or because the system is not observable - which implies the use of a state estimator). Even with a quasi-perfect model of the system, open-loop approaches can only yield suboptimal performances in unstructured stochastic environments, and can even lead to highly unstable behaviors if certain conditions are not met. Feedback control becomes then necessary to achieve the desired performances while adapting its strategies to tasks, environments or physical constraints. A goal-reaching trajectory may simply result from the feedback control laws. This explains the trial-to-trial variability of trajectories performed by humans during repetitive tasks, like hand motion in a goal-directed tasks: this variability cannot be explained by a controller performing trajectory tracking (i.e. if it tracks a pre-defined desired trajectory), but it is captured by a feedback controller that tries to reduce global task errors and makes the controlled trajectory robust to perturbations by varying impedance through co-contraction [50].

However, fast and coordinated limb movements cannot be executed under pure feedback control alone, because biological feedback loops are too slow (i.e. typical delays in the human sensory system are in the order of tenths of milliseconds or higher) and have small gains. Plausibly, a feedforward loop anticipates the evolution of the system and accounts for a desired predicted trajectory [41]. Together with feedback control relying on sensory measures, feedforward commands are employed to precompensate for the effects of actions, while forward models are used to predict these effects [177]. To perform this anticipation, the feedforward loop calls upon a state estimator and a forward model to predict the sensory consequences of actions based on motor commands [135]. Exploiting the so called "efference copy", the prediction can be used to refine the control strategies before the delayed sensory feedback to calibrate movements continuously and to improve the ability of the sensory system to estimate the state of the body and the environment. A deeper analysis of the possible implementations of such behaviors in the *CNS* would be far beyond the scope of this survey, see for example [139, 141] for details.

3.2. Which is the correct "cost function"?

Several experiments, focusing on the precomputation of feedback and feedforward motor commands [40, 41, 51, 169, 175], and their adaptation to changing environments [33, 71] have suggested "cues" of optimality principles in sensori-motor control. Meanwhile, the idea of optimization principles underlying *HMC* has been adopted in a va-

[2] For simple tasks, for example point-to-point arm movements, trajectories are overall stereotyped. However, when more freedom is given to subjects and tasks become more complex, more differences can arise [30]. These observations suggest that individual factors could play a role in the optimization and movement production process of the *CNS*. Put differently, individuals may optimize different costs or similar composite costs but with different weights when performing the same task [14].

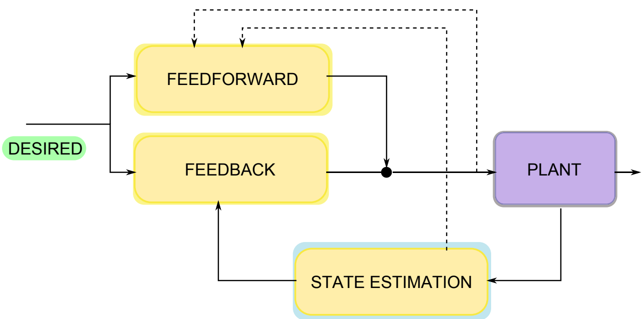

Figure 3. *Integration of feedback and feedforward commands in an optimal control scheme. Dashed arrows indicate possible flows of information, which can improve the feedforward controller if using internal models or efference copy. Remarkably, this conceptual scheme can be applied both for studying human motions and for implementing a controller for a humanoid robot.*

riety of studies, e.g. balance [21], walking [4, 99, 119], reaching [2, 86, 95, 109], goal-directed movements [54, 62, 172], adaptation to disturbances [16, 23]. During the last years, the analysis of the corresponding human motion data has provided a better understanding of *HMC* through computational models inspired by automatic controls, but also a variety of possible cost functions. If researchers tend to agree on the structure of computational models (e.g. a combination of feedback and feedforward terms), there is no consensus on the corresponding cost functions.

Moreover, this research is affected by the so called "observability problem". In general, to infer optimization principles out of human measured data an inverse problem is stated. It consists in determining the cost function which predicts the control trajectories observed in humans by means of an *OC* problem. The human recorded trajectories are the known optimal solution to the control problem, while the unknown is the optimized cost function [150]. The inverse problem is intrinsically ill-posed and it requires the exploration of an infinite-dimensional functional space and the solution of an inverse optimization problem, which is hard and impossible in most cases [46, 83]. However, by means of several simplifications, this approach can be used to "transfer" biological motions into robots. A pioneering work in this case was done by Mombaur et al. for human locomotion [100].

In brief, researchers trying to infer the optimization criteria out of human experimental data inevitably face the evidence that multiple cost functions lead to the same observed behavior. Consequently, models for explaining data suffer from similar issues: different models might give similar predictions on the same experimental data, and as long as contradictory experiences are not found, both models can be considered valid. Figure 4 illustrates the concept on different planning tasks involving the arm. Different cost functions (e.g. minimum hand jerk and minimum torque change) might give similar predictions on a task (e.g. point-to-point reaching, left side of Figure 4) and significantly different predictions on another (e.g. point-to-bar reaching, right side of Figure 4).

In a sense, it often happens that different models could explain the same behavior and that, despite the variety of principles proposed in the models, it is difficult to confute the soundness of one model against the others. Generally, the leading assumptions on the cost function are mainly inspired by cues emerging from human behaviors, experimental observation and researcher's intuition, which justifies the multitude of available models and the disputes in the *HMC* community.

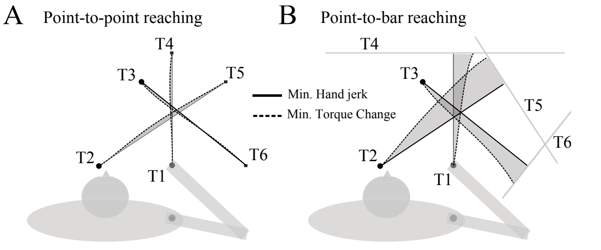

Figure 4. *A comparison of different costs on two different tasks: point-to-point reaching and point-to-bar reaching (image from [14]).*

Systematic approaches to the problem are still lacking in literature. The uniqueness of the cost function has only been recently addressed, for example in [14] for arm reaching movements. Weighted combinations of known cost functions, belonging to known parameterized classes have been proposed as an alternative to solving the inverse problem, which would result in looking one over infinitely many possible cost functions if no a priori assumptions are made on its structure.

In the following, we give an overview of the many cost functions which have been used to study point-to-point reaching movements. A summary of the criteria is reported in Table 1. Far from being comprehensive, we believe this list is a good example of the evolution of motor control models as seen from a robotics perspective.

Criterion	Cost function $\mathcal{J}$	References
Hand jerk	$\int_0^T \dddot{x}^2 + \dddot{y}^2 dt$	[49]
Angle jerk	$\int_0^T \dddot{\theta}_1^{\,2} + \dddot{\theta}_2^{\,2} dt$	[171]
Angle acceleration	$\int_0^T \ddot{\theta}_1^{\,2} + \ddot{\theta}_2^{\,2} dt$	[11]
Torque change	$\int_0^T \dot{\tau}_1^{\,2} + \dot{\tau}_2^{\,2} dt$	[163]
Torque	$\int_0^T \tau_1^2 + \tau_2^2 dt$	[108]
Geodesic	$\int_0^T [\dot{\theta}^\top \mathcal{M}(\theta)\dot{\theta}]^{1/2} dt$	[19]
Energy	$\int_0^T \left\lvert\dot{\theta}_1 \tau_1\right\rvert + \left\lvert\dot{\theta}_2 \tau_2\right\rvert dt$	[15]
Effort	$\int_0^T \mu_1^2 + \mu_2^2 dt$	[58]

Table 1. *Different cost functions (and related computational motor control models) for point-to-point movements. See Appendix B and the cited papers for the meaning of the variables.*

3.2.1. Minimum jerk

Based on the experimental evidence that goal-directed movements such as reaching or pointing result in straight hand paths with smooth velocity profiles, [49] proposed the *Minimum Jerk Model* (MJM) to describe the planar trajectories of the human arm while performing unconstrained point-to-point movements. The trajectories predicted by the MJM are straight-line Cartesian paths with bell-shaped velocity profiles, which is consistent with the experimental data for rapid human movements in the absence of accuracy constraints.

3.2.2. Minimum torque change

The main weakness of the MJM is that it always predicts straight paths, in contradiction with wide range movements and curved trajectories which occur for example during transverse movements, regardless of the influence of arm dynamics, arm posture, external forces, and movement duration.

Overcoming this issue, [163] proposed the *Minimum Torque Change Model* (MTCM), where trajectories are selected so as to minimize the rate of changes in torques. The MTCM takes into account the arm dynamics, and is able to reproduce gradually curved trajectories. Ten years later, [106] proposed a variant of MTCM, called The *Minimum Commanded Torque Change Model* (MCTCM), which provides a computable approximation of the MTCM while taking into account both link and muscle dynamics. The MTCM assumes null viscosity in the arm model, while MCTCM uses a non-null viscosity matrix in calculating the joints torques, thus considering both link dynamics and muscles as controlled objects in the model.

3.2.3. The Inactivation Principle

In 2008, [15] proposed a cost including a term called "absolute work of forces", reflecting the mechanical energy effort of a motion. In contrast to previous models, this term is non-smooth and non-differentiable, being based on an absolute function. According to the corresponding principle, supported by experimental observations from EMG signals, minimizing absolute terms implies simultaneous inactivation of agonistic and antagonistic muscles acting on a single joint, near the time of peak velocity.

3.2.4. Minimum endpoint variance and minimum intervention

In 1998, [61] observed that both eyes and arm movements are generated by neural controls corrupted by a signal-dependent noise, i.e. whose variance is proportional to the amount of control signal itself. Rapid motions, requiring larger control signals, would deviate from the desired trajectory as an effect of the disturbed control, resulting at the end in unsuccessful or imprecise final positions. Thus, they proposed the *Minimum Variance Theory* (MVT) which states that the accuracy in goal-directed movements is maximized by minimizing the variance of the final configuration.
The MVT is mostly an open-loop control principle. As control signal-dependent noise accumulates over the movement, a larger deviation at the endpoint of movement is observed with the increase of the signals. Consequently, in order to fulfill the minimum endpoint variance criterion, the motor system should activate the muscles as few as possible. These observations led to the Minimum Intervention Principle [159], that extends the MVT principle into an optimal feedback control loop. Furthermore, if the involved mechanical system is redundant with respect to the task, the motor control system should not react to perturbations that have no effect on the achievement of the task, leading to an "uncontrolled manifold" phenomenon [154].
The formalism behind the corresponding theory of motor coordination comes from *SOC*.

3.2.5. Risk sensitivity

Finally, [104] suggested that humans not only optimize the average cost associated to a movement, but being risk-sensitive, while optimizing the mean payoff they also take into account the variability of the payoff itself. In other words, they minimize the average cost together with its mean variance. These claims suggest that multi-objective optimization should be used to address the optimization problem behind *HMC*.

3.3. Optimality and movement duration

The time needed to perform a specific movement changes according to the circumstances, particularly in the presence of stochasticity or unforeseen perturbations, and generally depending on the precision of the movement. However, most computational models explaining human recorded data, for example during reaching, consider a fixed movement time (e.g. i*LQG* [160]), since the classical formulation of an *OC* problem usually requires the movement duration as a predefined constant.
It is then crucial to explain how the duration of a movement emerges from the control process instead of considering it as a predetermined input of that process. Most studies about movement duration are grounded on Fitts' and Schmidt' laws [48, 132], which relate the average movement duration to the amplitude of motion and to the error tolerance. The idea behind Fitts' law is that a certain amount of time is required to perform a movement, but the more precise a movement has to be (e.g. we want to touch a pin instead of a big ball) the more time is required to "adjust" the final position to the target. Several analysis and extensions to this law have been proposed, in particular for 2D tasks [91]. These and other models, such as the "minimum time principle" [149], predict movement duration correctly, but only for point-to-point movements. The MVT explains Fitts' law about the speed-accuracy trade-off, but still requires a movement duration to be pre-determined [57].
Recently, [140] and [123] proposed two similar explanations of movement duration based on the optimization of a trade-off between movement cost and a reward that is decayed through time. The latter proposes a model using an *OC* method for reaching based on a forward model. The cost function combines a cost term that is greater when the movement is faster, and a reward term that is greater if the goal is reached faster. The optimal time of movement emerges from the combination of both terms. Alternative explanations for the speed of movement start from a more descriptive approach, based on the decomposition of complex movements into units of motor action [85, 122, 166, 168]. Interestingly, these explanations are compatible with the optimality principles covered in this survey [12].

3.4. Adaptation in Motor Control

A constant adaptation mechanism is most likely entailed by the *HMC* system in order to cope with an evolving bio-mechanical system and a changing environment [140]. Motor adaptation is often interpreted as canceling the effects of novel environment on a noise rejection basis, so as to make the movements return to near baseline conditions. From a developmental perspective, it probably consists of a constant refinement of internal models and in a re-optimization process which computes a new optimal trajectory whenever an external perturbation is applied.
According to this view, stereotyped movement patterns are not prewired or inborn, but result from constant learning during ontogenesis. Infants dramatically improve their kinematic performance during their first months, but the developmental process towards stereotypical joint kinematics continues, as recently shown for locomotion [45] and earlier for reaching [81]. Trajectories straighten in time and the endpoint motion progressively gets smoother, although the unimodal velocity profiles and the inertial variability suggest that producing straight hand path may not be the most important criterion of the learning process [81]. Changes in the system structure or in the environment plausibly reflect into updates of the body schema and the internal models, which are exploited to re-optimize the trajectory planning and motor strategies. This continuous process has been observed when a new tool is used [29], and when obstacles impairing unconstrained reaching movements are introduced and removed from the workspace [41].
What happens to the optimization process when learning a new dynamic environment or when the environment changes? Current models claim that the subject performs at least two different computations: update the internal models (i.e. the mapping between the consequences of motor commands in terms of changes in the sensory states) and exploit the refined model to re-optimize the trajectory [71]. Alternatively,

the CNS could switch between different models, as suggested in [72]. In [176], perceived kinematic error was shown to play a role during adaptation, and subjects tended to maintain a visually straight path in front of perturbations. However, in [130], it was later shown that kinematic errors are not necessary for adaptation, i.e. the internal kinematic and dynamic model is continuously adapted even in the absence of visual feedback. In [97], it is shown that humans learn a new force field "dynamically" as opposed to solely rejecting the disturbances via increased stiffness and co-contraction. Once the internal forward model is properly learned, the CNS can re-optimize the motor cost, altering the baseline trajectory if necessary. Similarly, [71] suggests that adaptation entails accuracy and motor cost, and not the kinematic error from a desired baseline trajectory: thus, a re-optimization process computes a new optimal trajectory whenever an external perturbation is applied. Finally, another learning process related to habituation (known as "use-dependent learning") accounts for the tendency to reproduce similar joint trajectories along with the repetitions of a same task, though redundant [42].

In conclusion, motor adaptation entails both learning continuously accurate forward models, compensating for environmental changes, and finding the optimal controllers that maximize rewards / minimize costs of planned movements. When facing unpredictable tasks, like picking a box without knowing its load, the CNS initially generates highly variable behaviors, but eventually converges to stereotyped patterns of adaptive responses, which can be explained by simple optimality principles [23].

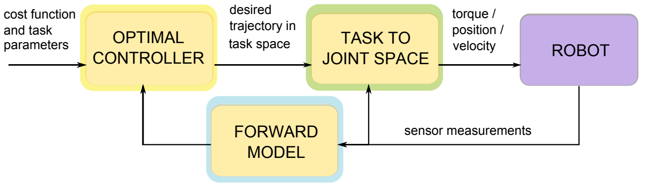

Figure 5. A conceptual scheme of a classic control scheme for robotics. The task parameters, such as the control function to be minimized, the current state of the robot, the task goal etc. are fed to the optimal planner, which computes the optimal trajectory, typically in the Cartesian space. An intermediate layer converts commands from operational (task) to joint space.

4. From humans to humanoids

Though the mechanisms of controlling and learning complex motor skills in humans are still largely unknown and vigorously debated, few general principles emerge that can account for the properties of human movements. The core of the $H2H$ approach is that, by implementing these principles on humanoid robots, it is possible not only to provide an experimental verification of the proposed models, but also to obtain behaviors able to outperform traditional classical controls [1, 6, 88]. In the second part of this survey, we investigate to what extent these HMC principles can also be used to generate the movement of robots and, in particular, humanoid robots.

4.1. Computational limitations

Many technical tools and notions used to model HMC, such as feed-forward and feedback loops (see Figure 3), come from automatic control and engineering sciences. However, the transfer of HMC principles to the control of robots (particularly using OC) is quite new.

A good explanation for this paradox comes from the high computational cost of OC methods, both in term of time and resources.

Notably, OC suffers from the *Curse of Dimensionality* (COD): the exponential growth of the number of parameters and in general of the computational complexity with the increase of the Degrees Of Freedom (DOF) of the problem. The time required by the algorithms to find the solution to the OC problem grows as well. This fact discriminates between robotics and neuroscience: indeed, in HMC studies there is no need to provide a solution to a control problem (e.g. a control trajectory) in real-time within an on-line control loop. To exemplify the computational cost of OC methods from a robotics point of view, in [162], a single movement generation for an HRP-2 robotic arm movement is reported to take from 1 to 4 minutes, even with a fast optimizer such as IPOPT [170]. Similarly, in [87], optimality is exploited to make kicking motions more accurate, exploiting a combination of an off-line planner aimed basically at minimizing torques, with a fast re-planning process, which adapts the controls depending on the current target configuration. However, though they also use IPOPT as an optimization tool, the authors admit that finding an optimal solution to an instance of the problem takes about two hours CPU time. State-of-the-art methods such as Nonlinear Model Predictive Control ($NMPC$) tools are still too expensive. Even the explicit precomputation of $NMPC$ laws is prohibitive for state/parameters spaces above $\mathbb{R}^{10}$ [44]. Even the solution to simplified problems (e.g., after strong hypotheses reducing the complexity of the model) cannot always guarantee the fulfillment of time constraints [44]. In contrast, a suitable (optimal) controller for a robot takes software and hardware limitations into account, e.g. complies to the control rate of the robot. Though local computations should be preferred because they could fasten the control cycle, it may not be feasible to perform such processing on local boards (i.e. the boards directly connected to the joints) if they have limited processing capabilities. A practical approach to this issue consists in dislocating the computations outside the robot, for instance on a cluster remotely connected to the robot. But in this configuration, real-time constraints cannot be guaranteed, and in general the safety of this control can be ensured only up to a certain level. Moreover, this generates important constraints on the potential context of use of the robot. Nevertheless, this strategy has been successfully applied in [9], where the optimization of a single trajectory for a 7 DOF arm is performed in real-time under numerous assumptions regarding the system dynamics and kinematics. Notably, most optimization is performed by a parallel computation on a cluster of 32 CPU cores, yielding 80% of success in the desired task.

In front of such heavy computational costs, there are different strategies to circumvent the resulting limitations, some purely practical or technical, some others being more grounded into theoretical properties of the control approaches. The remainder of the section is organized so as to cover the main research lines.

4.2. Analytical approaches

Thought numerical optimization is generally computationally expensive, there are cases when the analytical solution to OC problems is known [18, 126, 164]. For instance, under the well-known Linear Quadratic Gaussian (LQG) assumptions, explicit solutions can be found thanks to Riccati's equations. Researchers can thus profit of the numerous results from classical automatic control theory for their robotics applications. For example in [77], where a robust control is found combining a stabilizing control (based on Riccati and Lyapunov equations) and a neural network accounting for unknown dynamics; in [8] a robust neural sliding mode controller is presented, while tracking controllers are discussed in [22, 148].[3] Interestingly, among the different criteria reviewed in Section 3.2, an analytical solution exists and can be easily implemented for the MJM criterion. Since the closed-form solution is simple (a polynomial) and the implementation straightforward, it is frequently used in robotics. Furthermore, jerk minimization is beneficial if control strategies are implemented on real devices: since the velocity and acceleration profiles are very smooth, the system mechanics is less "stressed". An implementation of a MJM controller for the iCub humanoid robot can be found in [113].

Analytical solutions can be also used to tackle partially the control problem, in combination with numerical methods. For instance, in [77], a robust control is found combining a stabilizing control based on Riccati and Lyapunov equations and a neural network accounting for unknown dynamics. Another approach to using approximate analytical solutions for solving complex *OC* problems is presented in [89, 160]. It consists in starting from the resolution of a simple *LQR* (*LQG*) problem that approximates the original problem, and then iteratively refining the solution by adding a *LQR* (*LQG*) representation of the difference between what has been solved so far and the original problem. The resulting algorithm, called i*LQR* (in the deterministic case - i*LQG* in the Gaussian noise case) is among the state-of-the-art methods to implement *OC* in the context of *HMC* modeling as well as in transfer for robots [98]. i*LQG* has also been used in [16] to control the arm of the iCub robot during an active compliance task in presence of disturbances.

Unfortunately, in most cases analytical approaches are simply not feasible, and numerical approximations are solicited. For example in [43] *NMPC* with fast direct multiple shooting algorithm was used, and several approximations were made to reduce a 20 CPU seconds computations on a 3GHz Pentium IV to $200ms$, for a 5 state $150ms$ trajectory of a robotic arm. In [68] *NMPC* was combined with functional approximators to compute finite and receding horizon controllers for a planar arm.

4.3. Simplified models

A further impairment in the application of *OC* for the *H2H* transfer comes from the difficulty of adapting simple computational models to real robots. Indeed, most *HMC* models deal with simple simulated systems such as point-mass or 2 *DOF* planar arm (whereas a humanoid arm has usually 4 or 7 *DOF*-hands excluded- and moves unconstrained in its whole reachable 3D space). Nonetheless, these simplified models are of utmost interest because they allow researchers to circumvent the well-known issues of the *COD* by optimizing in a downsized space with a considerable saving of resources.

A classical approach to reduce the dimensionality is to transfer the control problem from a very large joint space (at the kinematics or dynamics level) or actuation space (which is larger than joint space when actuators are redundant, e.g. with muscles) to a generally much smaller task space. Task space trajectories can be then converted into suitable motor commands, taking into account the physical limitations of the platform, using precise models of the robot kinematics and dynamics (see Figure 5). For instance, if the task space is Cartesian, and joint velocity or position commands are used to control the robot motion, a classical Closed-Loop Inverse Kinematic (*CLIK*) loop can be used [28, 32]. This approach has been used for example in [69] with the humanoid James, and in [161], on a 39 *DOF* simulated humanoid robot.

More generally, researchers elude the *COD* by focusing on simplified motor control problems. The point-to-point reaching movement, where a 2 *DOF* arm is constrained to move in a plane, is a paradigmatic example at the heart of *HMC* research but also in many models for robotics [121]. Another strategy to tackle robotics control problems despite the *COD* consists in working with simplified models that "ignores" some *DOF* of the plant. This is particularly the case in locomotion studies, where the mechanical system can be often approximated with some success to a simple inverted pendulum [73].

A selection of significant examples for both approaches is presented below. Other examples that combine *OC* methods with adaptation capabilities, like [69], are studied in Section 4.5.

4.3.1. Reaching

In humanoid robotics, reaching is the fundamental action primitive. Rather than searching for generalized solutions to the reaching problem in the whole workspace, many approaches in literature focus on the optimization of single point-to-point movements [94, 137, 144, 162]. An implementation of the MVT for a 2 *DOF* arm was proposed in [144]. Models involving torques, such as the MCTCM, require the arm dynamics, thus a constrained nonlinear optimization problem must be solved, minimizing the cost function under hard constraints and boundary conditions.[4] In [74], a solution to the MCTCM is found by means of a numerical optimization of the Euler-Poisson equation: though describing a general procedure, the authors admit the impossibility to guarantee the convergence of the routine, thus making the algorithm unsuitable for real-time planning or control in robotic applications. In [142], *OC* is used to compute time-optimal motions of a robotic manipulator, considering nonlinear dynamics, actuator constraints, joint limits, and obstacles. In [179], an optimal motion planning problem is addressed to control a flexible space robot, in order to minimize the maneuvering time along with control and vibration energy. In [96], an *OC* problem is used to find controls for ball pitching with an under-actuated 2 *DOF* human-like arm, where in particular only the shoulder is actuated while the elbow is a passive spring with adaptive stiffness: the criterion is to maximize the ball velocity along a certain elevation angle. In [93, 95], the authors propose an experimentally-validated 3 *DOF* model of the human arm during constrained and unconstrained reaching movements. The cost to be minimized is based on energy and torque change, constrained by the hand-joint's freezing mechanism, explaining the experimental fact that the hand joint hardly changes its angle during reaching movements. Again *OC* theory is used to find the optimal trajectories of the hand during goal-directed motions.

4.3.2. Locomotion

Many researchers support the theory that optimization principles also explain the generation of gait and locomotion trajectories [84, 99]. In [153], the authors suggest that human walking analysis could improve the current humanoid robots walks, and particularly reduce the energy consumed during walking. In detail, they prove that a foot rotation subphase (specific during human fast walking) introduced in the gait contributes to the minimization of a torques-based cost, thus yielding optimal motions. In [4], the authors investigate human goal-directed

[3] Such classical control schemes are not suitable for cognitive functionalities, even if they come with a wide and assessed theory for stability, convergence and optimality. A combination of feedback and feedforward control laws can be used as a start (see Figure 3), but it is important to investigate new control strategies.

[4] The solution of this class of problems is generally difficult and, depending on the problem statement, there could be more than one method (or none) suited for its solution.

walking, assuming that locomotion trajectories are chosen according to some optimization principle. With the attempt to identify the optimized criteria (duration, length, etc.), they found that the time derivative of the curvature is minimized, and that trajectories are well-approximated by the geodesics minimizing the L_2 norm of the control, shaped as clothoids[5]. In [21], the authors suggest that the basic principle of animal locomotion is a mechanical rectification that converts periodic body movements to thrust force through interactions with the environment: thus, an optimal gait problem is formulated, where a quadratic cost function is minimized over a set of periodic functions subject to a velocity constraint, and the system is represented by a bilinear dynamic model, assuming small oscillations with respect to a nominal posture. In [64], *OC* is used for stable jumping of a one-legged hopping robot, with the goal to maximize energy efficiency of the motion. An interesting example of locomotion considering more *DOF* is presented in [134]: running is modeled as a multiphase periodic motion with discontinuities, based on multibody system models of the locomotion system with actuators and spring-damper elements at each joint; thus, running motions are generated offline as the solution of an *OC* problem, based on energy criteria, solved by an efficient direct multiple shooting algorithm. In [173], dynamic programming is used to optimize body motion, foot placement and step timing for a two link inverted pendulum model. Other examples of robotics implementation of *OC* methods for solving gait and locomotion problems can be found in [7, 31, 75, 90, 100].

4.4. Decomposition into primitives

Another possible approach to reduce the computational cost of finding the solution to *OC* problems consists in decomposing a global movement plan into the activation of a small set of motor primitives. This idea has profound roots in *HMC* since the pioneering ideas of Bernstein and the seminal work of [103], which reported evidence for linear combinations of motion primitives in the spinal cord of frogs. The appeal of decomposition into primitives comes from the associated dimensionality reduction: ideally, only a small finite number of scalar parameters should be set up to build an *OC* whenever appropriate primitives were stored in memory beforehand. Only relatively recently robot control studies have really attempted to expand upon this appealing concept. For instance, Nori and Frezza [110] built a mathematical framework of motion primitives to account for the presence of these linearly combinable spinal fields. They demonstrated the possibility to synthesize primitives from which a complete set of movements could be generated, and proved that controllability (as in control theory) could be preserved even when relying upon a small/finite set of modules. Crucial to their theory is the feedback linearization property of rigid body dynamics systems. Of course, the price to pay to exploit modularity in control may be a loss of optimality. However, recent advances in the context of *SOC* show that this is not always the case. Several investigators pointed out a specific, yet general, family of *SOC* problems for which the Bellman equation could be made linear [76]. Exploiting linearity, Todorov [157] exposed a theory of compositionality of control laws, based on the fact that task-optimal controllers can be constructed from certain primitives. Promising applications of this methodology to character animations have been reported for complex tasks such as diving, jumping and walking [34].

In the same vein, robotic studies on locomotion or similar rhythmic movements have been inspired by the presence of Central Pattern Generators (*CPGs*) in the human spinal cord. Schaal et al. [67, 128] developed dynamic models with autonomous nonlinear differential equations to create, in a flexible and modular way, smooth kinematic control policies. These dynamic movement primitives (*DMPs*) can then be mixed and tuned adequately to generate efficient behaviors with respect to any arbitrary cost function and with low computational load.

Altogether, these theoretical investigations suggest feasible ways of controlling complex stochastic nonlinear systems via a finite set of modules in an optimal or near-optimal way.

4.5. Machine Learning approaches

ML techniques can be easily integrated in the *OC* framework, with a twofold aim: first, to cope with evolving dynamics, and second, to tackle the *COD* through the use of incremental approaches. We cover both topics in the next two sections, and explain how they can be combined in a third section.

4.5.1. Adaptation to an evolving dynamics

Goal-directed movements are intended to reach a rewarding state at a minimum cost, but desired trajectories are not invariant with respect to system and environmental changes. In a life-long learning scenario, an active line of research consists in using on-line regression algorithms to learn or improve models of mechanical systems, at both kinematics and dynamics level. Within this approach, the internal models used to compute feedback and feedforward commands are constantly updated with the new information coming through the robot-environment interactions. Thus, they can account for changes at kinematic and dynamic level (e.g. a new tool at the end-effector, the suppression of a specific link in the kinematic chain, an obstacle impairing movement along one direction) but they can also be used to learn directly the forward/inverse models of the system when the system is complex [143].

More interestingly, *ML* can be combined with *OC* method for the production of controls. For example in [98], i*LQG* is combined with *LWPR* [165] to learn incrementally the model of a two-dimensional arm with 6 muscles, and reproduce the uncontrolled manifold phenomenon that comes with the minimal intervention principle presented in [159]. In [69] the Extended RItz Method is used to approximate finite and receding horizon controllers, which are optimizing the end-effector trajectory according to human plausible cost functions. The computational burden due to the optimization of the controllers, which are implemented as neural networks, is completely concentrated offline, while on-line controls are generated efficiently. The combination of the controllers with standard task space to joint space transformations (precisely the *CLIK* mentioned in Section 4.3) allows real-time control on the James humanoid robot, for reaching and tracking targets moving unpredictably.

4.5.2. Incremental computation of optimal controls

Whenever the solution of the *OC* problem does not comply with real-time requirements or is too greedy in terms of resources, *ML* techniques can be used to replace expensive computations with learnt controllers. In particular, the Reinforcement Learning (RL) community has been recently investigating the controller optimization problem (see [115] for an overview), where a parametric feedback controller, called *policy*, is improved over time through interactions of the robot with its environment. A parametrized controller can be learnt offline exploiting experimental datasets, incrementally improved from trial to trial online, or both. The main limit of these techniques is that learning is often stuck into local minima and that the correct initialization of the parameters is fundamental for the convergence of the algorithms. To circumvent these issues, it is a common practice to use a *Learning from*

[5] The clothoid or Cornu spiral is a curve, whose curvature grows with the distance from the origin.

Demonstration approach, which avoids learning from scratch and exploit a priori knowledge of the system or of the tasks (or both) [127]. The expected behavior is shown to the robot multiple times: these trials can be either used to learn incrementally the global controller or to compute an initial value for the parameters of the policy, so that the search is performed close enough to the desired optimum. This method is widely used in all the works listed below.

The first approach stems from the adaptation of discrete RL techniques to continuous state and action spaces. Early methods were based on Actor-Critic architectures, where an approximation of the expected performance of the policy is updated in parallel with the policy itself. Among these methods, Natural Actor-Critic (NAC) and its episodic variant [114] have been successfully applied to complex robotics problems [115]. However, they have been shown to be very difficult to tune [65] and unstable without adequate features.

Given these difficulties, the attention has shifted towards *direct Policy Search* methods (e.g. [25, 69, 79, 82]), which do not rely on an explicit representation of the expected performance. Instead, they optimize the parameters of a policy using stochastic optimization. The simplest of such methods is the Finite Differences method, which is based on the estimation of the gradient of the policy parameters with respect to the objective function by varying each dimension of the parameters in both directions. More powerful methods like *Episodic REINFORCE* [174], or the more recent developments inspired from Expectation-Maximization algorithms [79, 82], rely on an analytic derivation of the gradient of the objective function.

Probability-weighted averaging methods such as the 'Cross-Entropy Methods' (CEMs) [124], 'Covariance Matrix Adaptation - Evolutionary Strategy' (CMA-ES) [60] and 'Policy Improvement with Path Integrals' (PI^2) [25, 151] are even most robust, because they do not assume that the objective function is differentiable or even continuous and optimize a *population* of solutions rather than one (such as in gradient descent). CMA-ES can be seen as a variant of the CEM where the updates are "smoothed" over iterations (see [65] for a comparison with NAC). The more recent PI^2 algorithm is similar to CEM and CMA-ES, though it derives from different first principles [147]. It has been used for example to adapt the impedance of a movement over time [26, 146]. In [25], the parameters of a controller based on $DMPs$ were first tuned by showing the robot the required behavior, then were updated by the PI^2 algorithm performing a local search so as to optimize a cost function taking energy efficiency into account. One important advantage of PI^2 is that it is a model-free algorithm, which improves the performance of the controller by observing the cost function over a set of trajectories. Remarkably, the latter methods represent an interesting combination of ML and optimal motion planning.

Another instance of controller optimization with an incremental method is presented in [92]. A parametric controller for a planar arm with 6 muscles is learnt using the $XCSF$ [27] regression method from a set of optimal trajectories generated after solving a boundary problem as in [123]. Then a Cross-Entropy Policy Search ($CEPS$) algorithm (a policy search method similar to PI^2) is applied to the parametric controller, still using the same cost function as in [123]. As in [69], the advantage of the learnt controller is that it can be applied on-line without any computational burden (it is approximately 20000 times faster with respect to the on-line solution of a single OC problem).

4.5.3. Adaptation and re-optimization

The key property of incremental optimization is that the optimal response to a situation is not computed on the fly, but stored and retrieved when needed. One side effect of this process is that, when the circumstances are changing, the stored response may be suboptimal, or even inadequate, as mentioned in Section 2.2.

Nevertheless, the combination of incremental optimization with model adaptation is an important topic as it may provide a computational account of the phenomena investigated by [71] (see Section 3.4). To our knowledge, however, these topics have not yet been combined in a computational model, and likewise they do not have a counterpart in robotics. Therefore, they represent an interesting field of research to be further explored by combining ML, to address the learning problem, and SOC, to deal with differences between the real and learnt model finding optimal controls in unpredictable situations.

5. Conclusion and perspectives

In this paper, we have surveyed how OC principles extracted from HMC research can provide useful guidance in the design of advanced control solutions for robots. We focused on the use of OC principles for humanoid robotic control, and omitted other domains where the $H2H$ perspective may apply, such as the acquisition of sensorimotor transformations [118] and more generally, of spatial representations [3], diverse aspects of decision making under uncertainty [52, 152], the acquisition of behavioral schemata [131] and other cognitive phenomena. Being a very general framework, OC is ideal for transferring motor control principles from humans to humanoids. Particularly, a common formalism can be used to describe system models and cost functions, and can be easily adapted to the specific context of most human and humanoid control problems. Remarkably, SOC emerges as it is able to deal with uncertainties, and is naturally combined with ML techniques to provide adaptation capabilities.

However, there remains a theoretical and a practical limitation to the use of OC methods in the $H2H$ context. At the theoretical level, different cost functions have been proposed for explaining human movement data but yet the identification of the correct cost function suffers from the "observability problem". This issue is negligible as long as the different models are equally effective in reproducing human movement data; however, a choice among all possible models can be difficult for roboticists. Depending on the task and on the desired performances, one particular cost function could be preferred over the others, but others could be the criteria: ease of implementation, computational cost, and so on. In short, for the time being the choice of the cost function to use in robotics is left to the researcher. At the practical level, the main limitations to the application of these principles are related to the computational cost arising due to the COD. We have surveyed different approaches to circumvent the effects of the COD, and we have underlined the importance of ML methods to the development of the field. However, the implementation of such principles for complex tasks, like whole-body movements, is not straightforward.

There are already attempts of $H2H$ transfers exploiting such techniques, such as the pioneering works of [16, 25, 69], which combined incremental optimization of parametric policies, according to human-plausible costs, and realized implementations on robots.

Of course, a lot of progress is expected in the future from works that combine and improve several of the approaches described above. Beyond combination of these methods, a key challenge that has not been investigated yet lies in the modeling of broader architectures capable of integrating the achievement of several tasks so as to perform whole-body motion tasks in an optimal fashion. So far, HMC research has been mainly focused on elementary and paradigmatic phenomena such as motor adaptation in reaching or locomotion (using simplified models), rather than more integrated and complex motion problems. As a result, the neuroscience literature about the combined realization of several elementary tasks is rather sparse. The "compositionality" of OC laws is also relatively new [156]. But there is a growing tendency in

HMC science to address more complicated phenomena, which could bring insights into this problem. Recent works such as [53, 80] suggest that we may choose among several locally optimizing controllers based on the context of execution, rather than immediately adapting the current model to changes in the external environment.
More challenging to the OC view, the work of [66] suggests a less expensive solution to complex motion problems. The authors propose a model of accounting for the maximal end-state comfort effect that results from a simple combination of biases. Their model fits human trajectories as efficiently as OC models at a much lower computational cost.
Such solutions may result in computationally more tractable approaches to whole-body motion, but would still require a collection of controllers for achieving different tasks.
Finally, another line of research towards whole-body motion in humanoids consists in abstracting from the motor level some notions of actions that can be sequenced and combined into a more and more abstract hierarchy of options [102, 120]. The impact of motor costs in this hierarchical choice of actions remains to be studied.
Thus, as HMC research has stimulated robotics research on the production of elementary movements, we hope that, in the near future, robotics efforts towards whole-body motion in humanoids will stimulate HMC research towards a more integrated perspective on our movements.

Acknowledgments

We would like to thank the anonymous reviewers for suggestions that helped improving the paper. This work is supported by the French ANR program (ANR 2010 BLAN 0216 01, more at http://macsi.isir.upmc.fr) and in part by the European Projects: VIACTORS (FP7-ICT-2007-3), CHRIS (FP7- IST-215805) and ITALK (ICT-214668).

Appendix

A. Acronyms

Though mainly targeted to a robotic audience, this paper is also written for researchers from different fields, such as neuroscience. To help the reader in the variety of topics, we list hereby some of the most used acronyms on the paper.

CEMs	Cross-Entropy Methods
CMA-ES	Covariance Matrix Adaptation - Evolutionary Strategy
$CEPS$	Cross-Entropy Policy Search
$CLIK$	Closed Loop Inverse Kinematics
CNS	Central Nervous System
COD	Curse Of Dimensionality
$CPGs$	Central Pattern Generators
$DMPs$	Dynamic Motion Primitives
DOF	Degrees Of Freedom
EMG	Electro-Myo-Graphy
HMC	Human Motor Control
$H2H$	Humans To Humanoids
(i)LQG	(iterative) Linear Quadratic Gaussian
(i)LQR	(iterative) Linear Quadratic Regulator
$LWPR$	Locally Weighted Projection Regression
ML	Machine Learning
M(C)TCM	Minimum (Commanded) Torque Change Model
MJM	Minimum Jerk Model
MVT	Minimum Variance Theory
NAC	Natural Actor-Critic
$NMPC$	Nonlinear Model Predictive Control
PI^2	Policy Improvement through Path Integrals
RL	Reinforcement Learning
$(S)OC$	(Stochastic) Optimal Control

B. Optimal Control Problems

The goal of an OC problem is to provide the control signals that makes a system or a process reach its goal while satisfying its physical constraints and maximizing some performance criterion (or minimizing a cost, which will be our convention in this appendix). Here we focus on both deterministic and stochastic OC problems and aim to provide the reader with an informal presentation of the framework. The formulation of an OC problem usually requires a model, i.e. a mathematical description of the system, a list of physical constraints and the specification of the performance/cost function describing its task/behavior. In the control theory formalism, a deterministic system is described by a set of ordinary differential equations in state-space form as follows:

$$\dot{x}(t) = f(x(t), u(t), t) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is a set of variables fully describing the system at time t, $u(t) \in \mathbb{R}^m$ is the set of control inputs to the process at time t and $f(\cdot)$ is a function (possibly nonlinear) describing the evolution of the system. In the human or robot control literature, physical constraints are generally expressed as inequalities or equalities:

$$\begin{aligned} d(x(t), u(t), t) &\leq 0 \\ e(x(t), u(t), t) &= 0 \end{aligned} \tag{2}$$

The set of control sequences which satisfy the control constraints during the time horizon $[t_0, t_f]$ is called the set of admissible controls; whereas a state trajectory satisfying the state constraints (2) and (1) is called an admissible trajectory. The optimal control and system trajectories depend on the cost function, generally denoted by $\mathcal{J}$, which can take different forms. Generally, the Bolza form is considered:

$$\mathcal{J}(u(\cdot)) = h(x(t_f)) + \int_{t_0}^{t_f} g(x(t), u(t), t)dt$$

where h is the final/terminal cost and g is the immediate/running cost. We call optimal control the set of control sequences $u^\circ(t)$ which minimizes the functional $\mathcal{J}(u(\cdot))$ Then the optimal control problem can be formulated as follows: *find an admissible control $u^\circ(.)$, generating the admissible state trajectory $x^\circ(\cdot)$ for system (1), subject to constraints (2), such that the cost function $\mathcal{J}$ is minimized:*

$$u^\circ(\cdot) = \arg\min_{u(\cdot)} \mathcal{J}(u(\cdot)) \quad s.t. \quad (1)\,,\,(2)$$

There are many possible cost functions $\mathcal{J}$ that may be of interest depending on the task and the goals, as discussed in Section 3.2. The reader interested in the optimal control theory and in learning tools for solving such deterministic OC problems is referred to [117], [10], [78] or [24] for example.

Whenever the environment or the system itself is stochastic, the above theory has to be modified and adapted. Disturbances or uncertainties are usually modeled as random variables/processes. For example an additive noise can be modeled as a Wiener process (or Brownian motion), i.e. variables with specific probability distributions and properties. Then, the control problem is re-stated as a SOC problem. To do so, the deterministic system state (1) is then generally rewritten as a set of stochastic differential equations:

$$\dot{x}(t) = f(x(t), u(t), \eta(t), t) \tag{3}$$

where $\eta(t) \in \mathbb{R}^p$ is a generic set of stochastic variables affecting the system (typically a white noise). The rigorous meaning of (3) requires mathematics in the field of stochastic calculus (see for instance [111]). Note that $x(t)$ thus become a stochastic variable itself and has its own probability distribution over time depending on the control input $u(\cdot)$ (which can be either deterministic or stochastic depending on the problem formulation, e.g. open-loop vs feedback control laws). The cost function $\mathcal{J}$ becomes an *expected* cost function $J(u(\cdot))$ taken over all the realizations of the stochastic process $\eta(\cdot)$ (inducing different realizations of $x(\cdot)$ and/or $u(\cdot)$):

$$J(u(\cdot)) = E[\mathcal{J}(u(\cdot))] = E[h(x(t_f)) + \int_{t_0}^{t_f} g(x(t), u(t), t)dt]. \tag{4}$$

We may also consider constraints as in (2) (but with an expectation operator). A detailed analysis on stochastic dynamics and constraints is outside the scope of the paper, but the interested reader can refer to [18], [145] or [111] for SOC in general. Then a SOC problem can be formulated as follows: *find an admissible control u°, generating admissible trajectories $x^\circ(\cdot)$ for system (3), subject to constraints (2), such that the expected cost function J is minimized:*

$$u^\circ(\cdot) = \arg\min_{u(\cdot)} J(u(\cdot)) \quad s.t. \quad (3)\,,\,(2)$$

Note that the expectation operator $E[\cdot]$ is used, because the minimization of a stochastic cost is considered.

References

[1] Adams, B., Breazeal, C., Brooks, R. A., and Scassellati, B. (2000). Humanoid robots: a new kind of tool. *IEEE Intelligent Systems*, 15(4):25–31.

[2] Alexander, R. M. (1997). A minimum energy cost hypothesis for human arm trajectories. *Biological cybernetics*, 76(2):97–105.

[3] Andersen, R. A., Snyder, L. H., Bradley, D. C., and Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual review of Neuroscience*, 20(1):303–330.

[4] Arechavaleta, G., Laumond, J.-P., Hicheur, H., and Berthoz, A. (2008). An optimality principle governing human walking. *IEEE Transactions on Robotics*, 24:5–14.

[5] Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483.

[6] Atkeson, C. G., Hale, J. G., Pollick, F., Riley, M., Kotosaka, S., Schaul, S., Shibata, T., Tevatia, G., Ude, A., Vijayakumar, S., Kawato, E., and Kawato, M. (2000). Using humanoid robots to study human behavior. *IEEE Intelligent Systems and Their Applications*, 15(4):46–56.

[7] Atkeson, C. G. and Stephens, B. (2007). Multiple balance strategies from one optimization criterion. *2007 7th IEEE-RAS Int. Conf. on Humanoid Robots*, pages 57–64.

[8] Barambones, O. and Etxebarria, V. (2002). Robust neural control for robotic manipulators. *Automatica*, 38:235–242.

[9] Bauml, B., Wimbock, T., and Hirzinger, G. (2010). Kinematically optimal catching a flying ball with a hand-arm-system. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 2592–2599, Taipei, Taiwan.

[10] Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, Princeton, NJ.

[11] Ben-Itzhak, S. and Karniel, A. (2008). Minimum acceleration criterion with constraints implies bang-bang control as an underlying principle for optimal trajectories of arm reaching movements. *Neural Computation*, 20(3):779–812.

[12] Bennequin, D., Fuchs, R., Berthoz, A., and Flash, T. (2009). Movement timing and invariance arise from several geometries. *PLoS Computational Biology*, 5(7):e1000426.

[13] Bernstein, N. (1967). *The Co-ordination and Regulation of Movements*. Oxford, UK: Pergamo.

[14] Berret, B., Chiovetto, E., Nori, F., and Pozzo, T. (2011a). Evidence for composite cost functions in arm movement planning: An inverse optimal control approach. *PLoS Computational Biology*, 7(10):e1002183.

[15] Berret, B., Darlot, C., Jean, F., Pozzo, T., Papaxanthis, C., and Gauthier, J.-P. (2008). The inactivation principle: mathematical solutions minimizing the absolute work and biological implications for the planning of arm movements. *PLoS Computational Biology*, 4(10):e1000194.

[16] Berret, B., Ivaldi, S., Nori, F., and Sandini, G. (2011b). Stochastic optimal control with variable impedance manipulators in presence of uncertainties and delayed feedback. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 4354–4359.

[17] Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control*. Athena Scientific.

[18] Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-dynamic programming*. Athena Scientific.

[19] Biess, A., Liebermann, D. G., and Flash, T. (2007). A computational model for redundant human three-dimensional pointing movements: integration of independent spatial and temporal motor plans simplifies movement dynamics. *The Journal of Neuroscience*, 27(48):13045–13064.

[20] Billard, A., Calinon, S., Dillmann, R., and Schaal, S. (2007). *Handbook of Robotics (Siciliano, B. and Khatib, O. Eds)*, Robot Programming by Demonstration, pages 1371–1394. Springer.

[21] Blair, J. and Iwasaki, T. (2011). Optimal Gaits for Mechanical Rectifier Systems. *IEEE Transactions on Automatic Control*, 56(1):59–71.
[22] Braganza, D., Dixon, W. E., Dawson, D. M., and Xian, B. (2005). Tracking control for robot manipulators with kinematic and dynamic uncertainty. In *44th IEEE Conf. on Decision and Control*.
[23] Braun, D. A., Aertsen, A., Wolpert, D. M., and Mehring, C. (2009). Learning optimal adaptation strategies in unpredictable motor tasks. *The Journal of Neuroscience*, 29(20):6472–6478.
[24] Bryson, A. E. and Ho, Y.-C. (1975). *Applied Optimal Control: Optimization, Estimation, and Control*. John Wiley & Sons Inc.
[25] Buchli, J., Stulp, F., Theodorou, E., and Schaal, S. (2011). Learning variable impedance control. *The Int. Journal of Robotics Research*, 30(7):820–833.
[26] Buchli, J., Theodorou, E., Stulp, F., and Schaal, S. (2010). Variable impedance control - a reinforcement learning approach. In *Robotics Science and Systems*.
[27] Butz, M., Pedersen, G., and Stalph, P. (2009). Learning sensorimotor control structures with XCSF: redundancy exploitation and dynamic control. In *11th Annual Conf. on Genetic and Evolutionary Computation*, pages 1171–1178.
[28] Caccavale, F., Chiaverini, S., and Siciliano, B. (1997). Second-order kinematic control of robot manipulators with jacobian damped least-squares inverse: theory and experiments. *IEEE/ASME Transactions on Mechatronics*, 2(3):188–194.
[29] Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C., and Farnè, A. (2009). Tool-use induces morphological updating of the body schema. *Current Biology*, 19(12):R478–9.
[30] Cesqui, B., d'Avella, A., Portone, A., and Lacquaniti, F. (2012). Catching a ball at the right time and place: individual factors matter. *PLoS one*, 7(2):e31770.
[31] Chevallereau, C. and Aoustin, Y. (2001). Optimal reference trajectories for walking and running of a biped robot. *Robotica*, 19:557–569.
[32] Chiaverini, S., Egeland, O., and Kanestrom, R. K. (1991). Achieving user-defined accuracy with damped least-squares inverse kinematics. In *5th Int. Conf. on Advanced Robotics*, pages 672–677.
[33] Crevecoeur, F., Thonnard, J.-L., and Lefèvre, P. (2009). Optimal integration of gravity in trajectory planning of vertical pointing movements. *Journal of neurophysiology*, 102(2):786–796.
[34] da Silva, M., Durand, F., and Popovi¢, J. (2009). Linear Bellman combination for control of character animation. *ACM Transactions on Graphics*, 28(3):1.
[35] Dahiya, R. S., Metta, G., Valle, M., and Sandini, G. (2010). Tactile sensing: From humans to humanoids. *IEEE Transactions on Robotics*, 26(1):1–20.
[36] Dapena, J. (1980a). Mechanics of rotation in the fosbury-flop. *Medicine and Science in Sports and Exercise*, 12(1):45–53.
[37] Dapena, J. (1980b). Mechanics of translation in the fosbury-flop. *Medicine and Science in Sports and Exercise*, 12(1):37–44.
[38] Dapena, J. (2002). The evolution of high jumping technique: Biomechanical analysis. In *of 20th Internat. Symp. Biomech. Sports*, C ceres, Spain.
[39] De Santis, A., Siciliano, B., De Luca, A., and Bicchi, A. (2008). An atlas of physical human-robot interaction. *Mechanism and Machine Theory*, 43(3):253–270.
[40] Desmurget, M. and Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4:423–431.
[41] Diedrichsen, J., Shadmehr, R., and Ivry, R. B. (2010a). The coordination of movement: optimal feedback control and beyond. *Trends in Cognitive Sciences*, 14(1):31–39.
[42] Diedrichsen, J., White, O., Newman, D., and Lally, N. (2010b). Use-dependent and error-based learning of motor behaviors. *Journal of Neuroscience*, 30(15):5159–5166.
[43] Diehl, M., Bock, H. G., Diedam, H., and Wieber, P. B. (2006). *Fast Motions in Biomechanics and Robotics (Diehl, M. and Mombaur, K. Eds)*, vol. 340, Fast Direct Multiple Shooting algorithms for optimal robot control, pages 65–93. LNCIS, Springer.
[44] Diehl, M., Ferreau, H. J., and Haverbeke, N. (2009). *Nonlinear Model Predictive Control (Magni, L. et al. Eds)*, vol. 384, Efficient numerical methods for nonlinear MPC and Moving Horizon estimation, pages 541–550. LNCIS, Springer.
[45] Dominici, N., Ivanenko, Y. P., Cappellini, G., d'Avella, A., Mond , V., Cicchese, M., Fabiano, A., Silei, T., Di Paolo, A., Giannini, C., Poppele, R. E., and Lacquaniti, F. (2011). Locomotor primitives in newborn babies and their development. *Science*, 334(6058):997–999.
[46] Dupree, K., Patre, P., Johnson, M., and Dixon, W. (2009). Inverse optimal adaptive control of a nonlinear euler-lagrange system, part i: Full state feedback. In *48th IEEE Conf. on Decision and Control*, pages 321–326.
[47] Feldman, A. G. and Levin, M. F. (1995). The origin and use of positional frames of reference in motor control. *Behavioral and Brain Sciences*, 18(4):723–744.
[48] Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.*, 47(6):381–391.
[49] Flash, T. and Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, 5(7):1688–1703.
[50] Franklin, D. W., Burdet, E., Tee, K. P., Osu, R., Chew, C.-M., Milner, T. E., and Kawato, M. (2008). CNS learns stable, accurate, and efficient movements using a simple algorithm. *The Journal of Neuroscience*, 28(44):11165–11173.
[51] Franklin, D. W., So, U., Burdet, E., and Kawato, M. (2007). Visual feedback is not necessary for the learning of novel dynamics. *PloS one*, 2(12):e1336.
[52] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews*, 11:127–138.
[53] Ganesh, G., Albu-Schaffer, A., Haruno, M., Kawato, M., and Burdet, E. (2010). Biomimetic motor behavior for simultaneous adaptation of force, impedance and trajectory in interaction tasks. In *IEEE Int. Conf. on Robotics and Automation*, pages 2705–2711.
[54] Gepshtein, S., Seydell, A., and Trommershäuser, J. (2007). Optimality of human movement under natural variations of visual-motor uncertainty. *Journal of Vision*, 7(5):1–18.
[55] Gienger, M., Janssen, H., and Goerick, C. (2005). Task-oriented whole body motion for humanoid robots. In *5th IEEE-RAS Int. Conf. on Humanoid Robots*, pages 238–244.
[56] Guigon, E. (2011). *Motor Control (Danion, F. and Latash, M.L. Eds)*, Models and Architectures for motor control: Simple or complex?, pages 478–502. Oxford University Press.
[57] Guigon, E., Baraduc, P., and Desmurget, M. (2008a). Computational motor control: feedback and accuracy. *European Journal of Neuroscience*, 27(4):1003–1016.
[58] Guigon, E., Baraduc, P., and Desmurget, M. (2008b). Optimality, stochasticity and variability in motor behavior. *Journal of Computational Neuroscience*, 24(1):57–68.
[59] Haddadin, S., Albu-Schäffer, A., and Hirzinger, G. (2010). Safety analysis for a human-friendly manipulator. *Int. Journal of Social*

Robotics, 2:235–252.
[60] Hansen, N., Muller, S. D., and Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18.
[61] Harris, C. M. and Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394(6695):780–784.
[62] Harris, C. M. and Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biological Cybernetics*, 95(1):21–29.
[63] Hauser, H., Neumann, G., Ijspeert, A. J., and Maass, W. (2011). Biologically inspired kinematic synergies enable linear balance control of a humanoid robot. *Biological cybernetics*, 104(4-5):235–249.
[64] He, G.-P. and Geng, Z.-Y. (2007). Optimal motion planning of a one-legged hopping robot. In *IEEE Int. Conf. on Robotics and Biomimetics*, pages 1178–1183, Sanya, China.
[65] Heidrich-Meisner, V. and Igel, C. (2008). Similarities and differences between policy gradient methods and evolution strategies. In *16th Europ. Symp. on Artificial Neural Networks (ESANN)*, pages 149–154.
[66] Herbort, O. and Butz, M. (2011). The continuous end-state comfort effect: weighted integration of multiple biases. *Psychological Research*, pages 1–19.
[67] Ijspeert, A. J., Nakanishi, J., and Schaal, S. (2003). Learning attractor landscapes for learning motor primitives. In *Advances in Neural Information Processing Systems 15*, volume 15, pages 1547–1554.
[68] Ivaldi, S., Baglietto, M., Metta, G., and Zoppoli, R. (2009). *Non-linear Model Predictive Control (Magni, L. et al. Eds)*, vol. 384, An application of receding-horizon neural control in humanoid robotics, pages 541–550. LNCIS, Springer.
[69] Ivaldi, S., Fumagalli, M., Nori, F., Baglietto, M., and Metta, G. (2010). Approximate optimal control for reaching and trajectory planning in a humanoid robot. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 1290–1296, Taipei, Taiwan.
[70] Ivaldi, S., Fumagalli, M., Randazzo, M., Nori, F., Metta, G., and Sandini, G. (2011). Computing robot internal/external wrenches by means of inertial, tactile and F/T sensors: theory and implementation on the iCub. In *11th IEEE-RAS Int.Conf. on Humanoid Robots*, pages 521–528.
[71] Izawa, J., Rane, T., Donchin, O., and Shadmehr, R. (2008). Motor adaptation as a process of reoptimization. *The Journal of Neuroscience*, 28(11):2883–2891.
[72] Kadiallah, A., Liaw, G., Burdet, E., Kawato, M., and Franklin, D. W. (2008). Impedance control is tuned to multiple directions of movement. In *IEEE Int. Eng. Med. Biol. Soc. Conf.*, pages 5358–5361.
[73] Kajita, S., Kanehiro, F., Kaneko, K., Fujiwara, K., Harada, K., Yokoi, K., and Hirukawa, H. (2003). Resolved momentum control: Humanoid motion planning based on the linear and angular momentum. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, volume 2, pages 1644–1650.
[74] Kaneko, Y., Nakano, E., Osu, R., Wada, Y., and Kawato, M. (2005). Trajectory formation based on the minimum commanded torque change model using euler-poisson equation. *Systems and Computers in Japan*, 36:92–103.
[75] Kanoun, O., Yoshida, E., and Laumond, J.-P. (2009). An optimization formulation for footsteps planning. In *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 202–207, Paris, France.
[76] Kappen, H. J. (2005). A linear theory for control of non-linear stochastic systems. *Physical Review Letters*, 95:200–201.
[77] Kim, Y. H., Lewis, F. L., and Dawson, D. M. (2000). Intelligent optimal control of robotic manipulators using neural networks. *Automatica*, 36:1355–1364.
[78] Kirk, D. E. (1970). *Optimal control theory: An Introduction.* Prentice-Hall, New Jersey.
[79] Kober, J. and Peters, J. (2008). Policy search for motor primitives in robotics. *Advances in Neural Information Processing Systems (NIPS)*, pages 1–8.
[80] Kodl, J., Ganesh, G., and Burdet, E. (2011). The CNS Stochastically Selects Motor Plan Utilizing Extrinsic and Intrinsic Representations. *PLoS one*, 6(9):e24229.
[81] Konczak, J. and Dichgans, J. (1997). The development toward stereotypic arm kinematics during reaching in the first 3 years of life. *Experimental Brain Research*, 117:346–354.
[82] Kormushev, P., Calinon, S., and Caldwell, D. (2010). Robot motor skill coordination with em-based reinforcement learning. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 3232–3237.
[83] Krstic, M. (2009). Inverse optimal adaptive control : the interplay between update laws, control laws, and lyapunov functions. In *American Control Conf.*, pages 1250–1255.
[84] Kuo, A. (2005). An optimal state estimation model of sensory integration in human postural balance. *Journal of Neural Engineering*, 2:S235–S249.
[85] Lacquaniti, F., Terzuolo, C., and Viviani, P. (1983). The law relating kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54:115–130.
[86] Lan, N. and Crago, P. E. (1994). Optimal control of antagonistic muscle stiffness during voluntary movements. *Biological cybernetics*, 71(2):123–135.
[87] Lengagne, S., Ramdani, N., and Fraisse, P. (2009). Planning and fast re-planning of safe motions for humanoid robots: Application to a kicking motion. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 441–446.
[88] Lenzi, T., Vitiello, N., McIntyre, J., Roccella, S., and Carrozza, M. C. (2011). A robotic model to investigate human motor control. *Biological Cybernetics*, 105(1):1–19.
[89] Li, W. and Todorov, E. (2004). Iterative linear quadratic regulator applied to nonlinear biological movement systems. In *1st Int. Conf. on Informatics in Control, Automation and Robotics*, pages 222–229.
[90] Lockhart, D. and Ting, L. (2007). Optimal sensorimotor transformations for balance. *Nature Neuroscience*, 10(10):1329–1336.
[91] MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7:91–139.
[92] Marin, D. and Sigaud, O. (2012). Towards fast and adaptive optimal control policies for robots: A direct policy search approach. In *Proceedings Robotica*, pages 21–26.
[93] Matsui, T. (2008). A new optimal control model for reproducing two-point reaching movements of human three-joint arm with wrist joint's freezing mechanism. In *IEEE Int. Conf. on Robotics and Biomimetics*, pages 383–388.
[94] Matsui, T., Honda, M., and Nakazawa, N. (2006). A new optimal control model for reproducing human arm's two-point reaching movements: a modified minimum torque change model. In *IEEE Int. Conf. on Robotics and Biomimetics*, pages 1541–1546.
[95] Matsui, T., Takeshita, K., and Shibusawa, T. (2009). Effectiveness of human three-joint arm's optimal control model characterized by hand-joint's freezing mechanism in reproducing constrained reaching movement characteristics. In *ICROS-SICE*

Int. Joint Conf., pages 1206–1211.

[96] Mettin, U., Shiriaev, A. S., Freidovich, L. B., and Sampei, M. (2010). Optimal ball pitching with an underactuated model of a human arm. In *IEEE Int. Conf. on Robotics and Automation*, pages 5009–5014.

[97] Mistry, M., Theodorou, E., Liaw, G., Yoshioka, T., Schaal, S., and Kawato, M. (2008). Adaptation to a sub-optimal desired trajectory. In *Society for Neuroscience - Symp. on Advances in Computational Motor Control*, Washington DC, USA.

[98] Mitrovic, D., Klanke, S., and Vijayakumar, S. (2010). *From Motor to Interaction Learning in Robotics (Sigaud, O. and Peters, J. Eds)*, vol. 264, Adaptive Optimal Feedback Control with Learned Internal Dynamics Models, pages 65–84. Springer-Verlag.

[99] Mombaur, K., Laumond, J.-P., and Yoshida, E. (2008). An optimal control model unifying holonomic and nonholonomic walking. In *8th IEEE-RAS Int. Conf. on Humanoid Robots*, Daejeon, Korea.

[100] Mombaur, K., Truong, A., and Laumond, J.-P. (2010). From human to humanoid locomotion: an inverse optimal control approach. *Autonomous Robots*, 28:369–383.

[101] Morasso, P. (1983). Three dimensional arm trajectories. *Biological Cybernetics*, 48:187–194.

[102] Mugan, J. and Kuipers, B. (2009). Autonomously learning an action hierarchy using a learned qualitative state representation. In *Int. Joint Conf. on Artificial Intelligence*.

[103] Mussa-Ivaldi, F. A., Giszter, S. F., and Bizzi, E. (1994). Linear combinations of primitives in vertebrate motor control. *Proc. National Academy of Sciences USA*, 91(16):7534–7538.

[104] Nagengast, A. J., Braun, D. A., and Wolpert, D. M. (2011). Risk-sensitivity and the mean-variance trade-off: decision making in sensorimotor control. *Proceedings Biological Sciences / The Royal Society*, 278(1716):2325–2332.

[105] Nakamura, Y. (1991). *Advanced Robotics: redundancy and optimization*. Addison Wesley.

[106] Nakano, E., Imamizu, H., Osu, R., Uno, Y., Gomi, H., Yoshioka, T., and Kawato, M. (1999). Quantitative examinations of internal representations for arm trajectory planning: Minimum commanded torque change model. *Journal of Neurophysiology*, 81:2140–2155.

[107] Nakaoka, S., Nakazawa, A., Yokoi, K., Hirukawa, H., and Ikeuchi, K. (2003). Generating whole body motions for a biped humanoid robot from captured human dances. In *IEEE Int. Conf. on Robotics and Automation*, volume 3, pages 3905–3910.

[108] Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, 46:135–147.

[109] Nishii, J. and Murakami, T. (2002). Energetic optimality of arm trajectory. In *Int. Conf. on Biomechanics of Man*, pages 30–33.

[110] Nori, F. and Frezza, R. (2005). A control theory approach to the analysis and synthesis of the experimentally observed motion primitives. *Biological Cybernetics*, 93(5):323–342.

[111] Oksendal, B. (1995). *Stochastic Differential Equations*. Springer Berlin, 4th edition.

[112] Parker, G. A. and Smith, J. M. (1990). Optimality theory in evolutionary biology. *Nature*, 348(6296):27–33.

[113] Pattacini, U., Nori, F., Natale, L., Metta, G., and Sandini, G. (2010). An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Taipei, Taiwan.

[114] Peters, J. and Schaal, S. (2007). Natural actor-critic. *Neurocomputing*, 71:1180–1190.

[115] Peters, J. and Schaal, S. (2008). Reinforcement learning of motor skills with policy gradients. *Neural networks*, 21:682–97.

[116] Pollard, N. S., Hodgins, J. K., Riley, M. J., and Atkeson, C. G. (2002). Adapting human motion for the control of a humanoid robot. In *IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 1390–1397.

[117] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishchenko, E. F. (1964). *The Mathematical Theory of Optimal Processes*. Pergamon Press.

[118] Pouget, A. and Snyder, L. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3:1192–1198.

[119] Pozzo, T., Berthoz, A., and Lefort, L. (1990). Head stabilisation during various locomotor tasks in humans. i. normal subjects. *Experimental Brain Research*, 82:97–106.

[120] Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., and Botvinick, M. M. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, 71(2):370–379.

[121] Richardson, M. J. E. and Flash, T. (2000). On the emulation of natural movements by humanoid robots. In *IEEE-RAS Int. Conf. on Humanoids Robots*.

[122] Richardson, M. J. E. and Flash, T. (2002). Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. *Journal of Neuroscience*, 22(18):8201–8211.

[123] Rigoux, L., Sigaud, O., Terekhov, A., and Guigon, E. (2010). Movement duration as an emergent property of reward directed motor control. In *Annual Symp. Advances in Computational Motor Control*.

[124] Rubinstein, R. Y. (1997). Optimization of computer simulation models with rare events. *European Journal of Operational Research*, 99(1):89–112.

[125] Salini, J., Padois, V., and Bidaud, P. (2011). Synthesis of complex humanoid whole-body behavior: a focus on sequencing and tasks transitions. In *IEEE Int. Conf. on Robotics and Automation*, pages 1283–1290.

[126] Sastry, S. and Bodson, M. (1994). *Adaptive Control: Stability, Convergence, and Robustness*. Advanced Reference Series (Engineering). Prentice-Hall.

[127] Schaal, S. (1997). *Advances in Neural Information Processing Systems (Mozer, M.C. et al. Eds)*, Learning from demonstration, pages 1040–1046. MIT Press.

[128] Schaal, S., Peters, J., Nakanishi, J., and Ijspeert, A. J. (2003). Learning movement primitives. In *Int. Symp. on Robotics Research (ISRR)*, pages 561–572.

[129] Schaal, S. and Schweighofer, N. (2005). Computational motor control in humans and robots. *Current Opinion in Neurobiology*, 15:675–682.

[130] Scheidt, R. A., Reinkensmeyer, D. J., Conditt, M. A., Rymer, W. Z., and Mussa-Ivaldi, F. A. (2000). Persistence of motor adaptation during constrained, multi-joint, arm movements. *Journal of Neurophysiology*, 84(2):853–862.

[131] Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological review*, 82(4):225.

[132] Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., and Quinn, J. T. (1979). Motor output variability: a theory for the accuracy of rapid motor acts. *Psychol. Rev.*, 47:415 51.

[133] Schöner, G. and Kelso, J. A. (1988). Dynamic pattern generation in behavioral and neural systems. *Science*, 239(4847):1513.

[134] Schultz, G. and Mombaur, K. (2010). Modeling and optimal control of human-like running. *IEEE/ASME Trans. on Mechatronics*, 15(5):783–792.

[135] Scott, S. (2004). Optimal feedback control and the neural basis

of volitional motor control. *Nature Reviews Neuroscience*, 5:532–546.

[136] Scott Kelso, J. A. (1982). *Human motor behavior: an introduction*. Lawrence Erlbaum Associates.

[137] Seki, H. and Tadakuma, S. (2004). Minimum jerk control of power assisting robot based on human arm behavior characteristics. In *IEEE Int. Conf. on System, Man and Cybernetics*, volume 1, pages 722–727.

[138] Sentis, L. and Khatib, O. (2005). Synthesis of whole-body behaviors through hierarchical control of behavioral primitives. *The Int. Journal of Humanoid Robotics*, 2(4):505–518.

[139] Shadmehr, R. and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185:359–381.

[140] Shadmehr, R., Orban de Xivry, J.-J., Xu-Wilson, M., and Shih, T.-Y. (2010). Temporal discounting of reward and the cost of time in motor control. *The Journal of Neuroscience*, 30(31):10507–10516.

[141] Shadmehr, R. and Wise, S. (2005). *The Computational Neurobiology of Reaching and Pointing: a foundation for Motor Learning*. MIT Press.

[142] Shiller, Z. and Dubowsky, S. (1991). On computing the global time-optimal motions of robotic manipulators in the presence of obstacles. *IEEE Transactions on Robotics and Automation*, 7(6):785–797.

[143] Sigaud, O., Salaün, C., and Padois, V. (2011). On-line regression algorithms for learning mechanical models of robots: a survey. *Robotics and Autonomous Systems*, 51:1117–1125.

[144] Simmons, G. and Demiris, Y. (2005). Optimal robot arm control using the minimum variance model. *Journal of Robotic Systems*, 22(11):677–690.

[145] Stengel, R. F. (1994). *Optimal Control and Estimation*. Dover Publications.

[146] Stulp, F., Buchli, J., Ellmer, A., Mistry, M., Theodorou, E., and Schaal, S. (2011). Reinforcement learning of impedance control in stochastic force fields. In *IEEE Int. Conf. on Development and Learning*, volume 2, pages 1–6.

[147] Stulp, F. and Sigaud, O. (2012). Path integral policy improvement with covariance matrix adaptation. In *29th Int. Conf. on Machine Learning*.

[148] Sun, F. C., Li, H. X., and Li, L. (2002). Robot discrete adaptive control based on dynamic inversion using dynamical neural networks. *Automatica*, 38:1977–1983.

[149] Tanaka, H., Krakauer, J. W., and Qian, N. (2006). An optimization principle for determining movement duration. *Journal of Neurophysiology*, 95:38750–3886.

[150] Terekhov, A. V. and Zatsiorsky, V. M. (2011). Analytical and numerical analysis of inverse optimization problems: conditions of uniqueness and computational methods. *Biological Cybernetics*, 104:75–93.

[151] Theodorou, E., Buchli, J., and Schaal, S. (2010). Reinforcement learning of motor skills in high dimensions: a path integral approach. In *Int. Conf. on Robotics and Automation*, pages 2397–2403. IEEE.

[152] Thrommershäuser, J., Maloney, L. T., and Landy, M. S. (2008). Decision making, movement planning, and statistical decision theory. *Trends in Cognitive Sciences*, 12(8):291–297.

[153] Tlalolini, D., Chevallereau, C., and Aoustin, Y. (2011). Human-like walking: Optimal motion of a bipedal robot with toe-rotation motion. *IEEE/ASME Transactions on Mechatronics*, 16(2):310–320.

[154] Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9):907–915.

[155] Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural computation*, 17(5):1084–1108.

[156] Todorov, E. (2009a). Compositionality of optimal control laws. *Advances in Neural Information Processing Systems*, 3:2–6.

[157] Todorov, E. (2009b). Efficient computation of optimal actions. *Proc. Natl. Acad. Sci. USA*, 106(28):11478–11483.

[158] Todorov, E. and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235.

[159] Todorov, E. and Jordan, M. I. (2003). A minimal intervention principle for coordinated movement. In *Advances in Neural Information Processing Systems*, volume 15, pages 27–34.

[160] Todorov, E. and Li, W. (2005). A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *American Control Conf.*, pages 300–306.

[161] Toussaint, M., Gienger, M., and Goerick, C. (2007). Optimization of sequential attractor-based movement for compact behaviour generation. In *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 122–129.

[162] Tuan, T., Souères, P., Taïx, M., and Guigon, E. (2008). A principled approach to biological motor control for generating humanoid robot reaching movements. In *IEEE Int. Conf. Biomedical Robotics and Biomechatronics*, pages 783–788.

[163] Uno, Y., Kawato, M., and Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, 61:89–101.

[164] Vidyasagar, M. (1987). *Control System Synthesis: A factorization approach*. MIT Press.

[165] Vijayakumar, S. and Schaal, S. (2000). Locally Weighted Projection Regression: An O(n) Algorithm for Incremental Real Time Learning in High Dimensional Space. In *7th Int. Conf. on Machine Learning*, pages 1079–1086.

[166] Viviani, P. (1986). *Generation and modulation of action patterns (Heuer, H. and Fromm, C. Eds)*, Do units of motor action really exist?, pages 201–216. Springer-Verlag.

[167] Viviani, P. and Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance*, 21:32–53.

[168] Viviani, P. and Stucchi, N. (1992). Biological movements look constant: Evidence of motor perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 18:603–623.

[169] Volkinshtein, D. and Meir, R. (2011). Delayed feedback control requires an internal forward model. *Biological cybernetics*, 105(1):41–53.

[170] Wächter, A. and Biegler, L. (2006). On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106:25–57.

[171] Wada, Y., Kaneko, Y., Nakano, E., Osu, R., and Kawato, M. (2001). Quantitative examinations for multi joint arm trajectory planning-using a robust calculation algorithm of the minimum commanded torque change trajectory. *Neural Networks*, 14(4-5):381–393.

[172] Wada, Y., Yamanaka, K., Soga, Y., Tsuyuki, K., and Kawato, M. (2006). Can a kinetic optimization criterion predict both arm trajectory and final arm posture? *Annual Int. Conf. of the IEEE Eng. Med. Biol. Society*, 1:1197–200.

[173] Whitman, E. C. and Atkeson, C. G. (2009). Control of a walking

biped using a combination of simple policies. In *IEEE-RAS Int. Conf. on Humanoid Robots*, pages 520–527, Paris, France.

[174] Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4):229–256.

[175] Wolpert, D. M. and Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 18(18):729–732.

[176] Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). Are arm trajectories planned in kinematic or dynamic coordinates? an adaptation study. *Experimental Brain Research*, 103:460–470.

[177] Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9):338–347.

[178] Yoshida, E., Esteves, C., Kanoun, O., Poirier, M., Mallet, A., Laumond, J.-P., and Yokoi, K. (2010). *Motion Planning for Humanoid Robots (Harada, K. et al. Eds)*, Planning Whole-body Humanoid Locomotion, Reaching, and Manipulation, pages 99–128. Springer.

[179] Zhao, H. and Chen, D. (1996). Optimal motion planning for flexible space robots. In *IEEE Int. Conf. on Robotics and Automation*, pages 393–398.