The Interaction of Maturational Constraints and Intrinsic Motivations in Active Motor Development

Adrien Baranes, Pierre-Yves Oudeyer

The Interaction of Maturational Constraints and Intrinsic Motivations in Active Motor Development

Adrien Baranes and Pierre-Yves Oudeyer
INRIA, France

***Abstract*—This paper studies computational models of the coupling of intrinsic motivations and physiological maturational constraints, and argues that both mechanisms may have complex bidirectional interactions allowing the active control of the growth of complexity in motor development which directs an efficient learning and exploration process. First, we outline the Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity algorithm (SAGG-RIAC) that instantiates an intrinsically motivated goal exploration mechanism for motor learning of inverse models. Then, we introduce a functional model of maturational constraints inspired by the myelination process in humans, and show how it can be coupled with the SAGG-RIAC algorithm, forming a new system called McSAGG-RIAC^2. We then present experiments to evaluate qualitative and, more importantly, quantitative properties of these systems when applied to a 12DOF quadruped controlled with 24 dimensions motor synergies.**

I. Constrained Learning Process

The learning capabilities given to biological systems are the result of evolution. They allow the discovery and scaffolding of new skills whose usefulness cannot be biologically anticipated. As a process which begins in the first instants of life, learning is guided by several intrinsic and extrinsic mechanisms and is highly constrained by the evolving body plan of infants and animals, as well as the increasing processing capabilities of their brain. Several questions have been raised by psychologists and neuroscientists about the learning process and the formation of new know-how, but only a few explored the complex bidirectional interactions between learning and biological evolution. Infancy is indeed the period where the maximum amount of information and know-how is learned, but also when the human body evolves the most significantly.

The study proposed in this paper aims to show how complex bidirectional interactions between the learning process and constraints could produce the efficient progressive learning of numerous skills inside unbounded high-dimensional spaces. We propose a qualitative and quantitative analysis of the learning efficiency of McSAGG-RIAC^2, an updated version of the McSAGG-RIAC (Maturationally-Constrained Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity) algorithm that we previously introduced in [1] and studied only in a qualitative manner. We introduce this mechanism in the case of a 12DOF quadruped controlled with motor synergies specified in 24 dimensions, which learns to reach different positions, and evaluate the high potential of McSAGG-RIAC^2 to efficiently guide the learning process.

A. Intrinsic Motivation Systems

Intrinsic motivations systems have been shown as a potential way to allow developmental robots to learn and discover new skills autonomously and in an incremental manner [2], [3]. Studied as active learning algorithms [4], [5], they have been shown as directing the emergence of developmental trajectories allowing an efficient organized and constrained self-exploration process. These mechanisms typically use meta-models of performance of the learning process to guide the exploration in the most "interesting" sensory-motor areas which maximize the notion of informational gain (e.g. heuristics for maximizing entropy, uncertainty, variance, prediction errors or decrease in prediction errors) [3], [4], [6]. While numerous heuristics are not robust to drive exploration in spaces where some parts cannot be learned or where the noise is inhomogeneous [7], some of them (e.g. decrease in prediction errors) have been shown to be efficient in such spaces [3], [8]–[10]

Nevertheless, all of these methods begin by a random and sparse exploration of the whole space to discriminate areas of different interests, which is an issue when considering learning in unbounded spaces where typical developmental robots evolve. This problem has been studied and partially resolved by competence based intrinsic motivation mechanisms [11], [12], such as the Self-Adaptive Goal Generation Robust-Intelligent Adaptive Curiosity (SAGG-RIAC) algorithm which considers as interesting the local improvement of its competence to reach high-level self-generated goals [13]. It also addresses an issue typically existing in intrinsically motivated algorithms such as the consideration of numerous ways to perform the same task instead of a single way to perform different tasks, by introducing high-level goals explicitly and driving exploration at their level.

B. Biological Constraints

Biological constraints represent all kinds of internal aspect of an embodiment which limits its access and interactions with the environment [14]. They can be explained by both the evolving structure of the brain of a learning agent as well as its body, which have strong influences on the way it perceives, acts, and interpret the environment where it evolves, hence its learning process [15]. Depending on the characteristics of its sensors, actuators and brain, an embodied system can be assisted, limited and constrained in its learning process from two different points of view: first, morphological and computational limitations of an embodied system can be seen as a way to delimit the environment accessible by an agent, and reduce the size of its explorable sensorimotor space [16], [17]. Second, prewired mechanisms present in the brain as well as the structure of the embodiment by itself can allow simplifications of both analysis of sensory data and control of motor actions [18].

C. Toward a Co-Optimization and Bidirectional Relations between Evolving Embodiment and Control

A growing number of studies began working on demonstrating how and why morphology and control should be coupled

(e.g., [19]–[21]). In this paper we argue that intertwining the different kinds of constraints presented above can serve as a basis for an open-ended learning framework. In the following sections, we will first present the Self-Adaptive Goal Generation RIAC algorithm (SAGG-RIAC) introduced in [13] as an original approach to Competence Based Active Motor Learning [11], [22], [23]. Then, we aim to show how intertwining driving mechanisms and constraints can allow efficient learning inside open-ended spaces of a maximum amount of self-generated skills. Practically, we merge maturational constraints and competence-based intrinsic motivations and present the algorithm **McSAGG-RIAC**^{2} which uses bidirectional interactions between these two processes in order to carry out the continuous active control of the release of constraints.

II. Competence Based Intrinsic Motivation: The Self-Adaptive Goal Generation RIAC Algorithm

A. Global Architecture

Let us consider the definition of competence based models outlined in [22], and extract from it two different levels for active learning defined at different time scales (see Fig. 1):

1) The higher level of active learning (higher time scale) considers the *active self-generation and self-selection of goals*, depending on feedback defined using the level of achievement of previously generated goals.
2) The lower level of active learning (lower time scale) considers the *goal-directed active choice and active exploration* of lower-level actions to be taken to reach the goals selected at the higher level, and depending on local measures of the evolution of the quality of learned inverse and/or forward models.

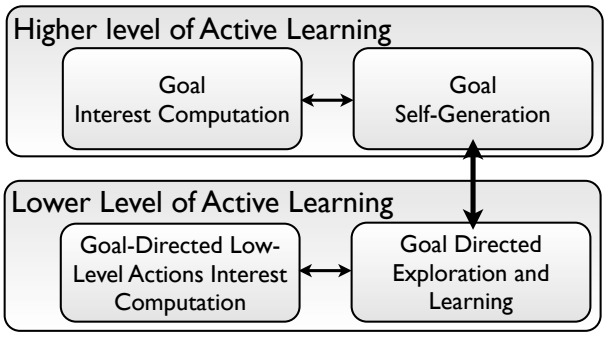

Fig. 1. Global Architecture of the SAGG-RIAC algorithm.

B. Model Formalization

Let us consider a robotic system whose configurations/states are described in both an actuator space S, and an operational/task space S'. For given configurations $(s_1, s'_1) \in S \times S'$, a sequence of actions $a = \{a_1, a_2, ..., a_n\}$ allows a transition towards the new states $(s_2, s'_2) \in S \times S'$ such that $(s_1, s'_1, a) \Rightarrow (s_2, s'_2)$. For instance, in the case of a robotic manipulator, S may represent its actuator/joint space, S' the operational space corresponding to the cartesian position of its end-effector, and a may be velocity or torque commands in the joints.

SAGG-RIAC considers the reaching of *goals* from starting states. Starting states are formalized as configurations $(s_{start}, s'_{start}) \in S \times S'$ and goals as a desired $s'_g \in S'$. All states are considered to be potential starting states; therefore, once a goal has been generated, the lower level of active learning is always able to try to reach it by starting from the current state of the system.

C. Lower Time Scale: Active Goal Directed Exploration and Learning

The goal directed exploration and learning mechanism can be carried out in numerous ways. Its main idea is to guide the system toward the goal by executing low-level actions which allow progressive exploration of the world and create a model that may be reused afterwards. Its implementation has to respect two imperatives :

1) A model (inverse and/or forward) has to be computed during exploration and has to be available for a later reuse, in particular when considering other goals.
2) A learning feedback mechanism has to be added such that the exploration is active, and the selection of new actions depends on local measures about the quality of the learned model.

In the following experiment that will be introduced, we will use an optimization algorithm. Other kinds of techniques, for example ones based on natural actor-critic architectures in model based reinforcement learning [24] or coming from evolutionary robotics [25], could also be used.

D. Higher Time Scale: Goal Self-Generation and Self-Selection

The Goal Self-Generation and Self-Selection process relies on feedback defined using the concept of competence, and more precisely on the competence improvement in given regions (or subspaces) of the space where goals are chosen. The measure of competence can be computed at different instants of the learning process. First, it can be estimated once a reaching attempt in direction of a goal has been declared as terminated. Second, for robotic setups which are compatible with this option, competence can be computed during low-level reaching attempts. In the following, we detail these two different cases.

1) Measure of Competence: We introduce a measure of competence for a given goal reaching attempt as dependent on two metrics: the similarity between the state s'_f attained when the reaching attempt has terminated, and the actual goal s'_g; and the respect of requirements ρ. These conditions are represented by the function of similarity Sim defined in $[-\infty; 0]$, such that the higher the $Sim(s'_g, s'_f, \rho)$ will be, the more a reaching attempt will be considered as efficient. From this definition, we set a measure of competence $\gamma_{s'_g}$ directly linked with the value of $Sim(s'_g, s'_f, \rho)$:

$$\gamma_{s'_g} = \begin{cases} Sim(s'_g, s'_f, \rho) & \text{if } Sim(s'_g, s'_f, \rho) \leq \varepsilon_{sim} < 0 \\ 0 & \text{otherwise} \end{cases}$$

where ε_{sim} is a tolerance factor and $Sim(s'_g, s'_f, \rho) > \varepsilon_{sim}$ corresponds to a goal reached. We note that a high value of $\gamma_{s'_g}$ (i.e. close to 0) represents a system that is competent in reaching the goal s'_g while respecting requirements ρ. A typical instantiation of Sim, without requirements, is defined as $Sim(s'_g, s'_f, \emptyset) = -\|s'_g - s'_f\|^2$, and is the direct transposition of prediction error in RIAC (which here refers to goal reaching error [3], [10]).

Measuring competence during reaching attempts thus allows taking advantage of every actions performed in order to compute a measure of competence, and improves the quantity of feedback send to the Goal Self-Generation mechanism.

2) Definition of Local Competence Progress: Let us consider different measures of competence $\gamma_{s'_i}$ computed for different attempted goals $s'_i \in S', i > 1$. For a subspace called a region $R \subset S'$, we can compute a measure of competence γ'' that we call a *local measure* such that:

$$\gamma'' = \left(\frac{\sum_{s'_j \in R} (\gamma_{s'_j})}{|R|} \right), \text{ with } |R|, \text{ cardinal of } R$$

Let us now consider different regions R_i of S' such that $R_i \subset S'$, $\bigcup_i R_i = S'$ (initially, there is only one region which is then progressively and recursively split; see below). Each R_i contains attempted goals $\{s'_{t_1}, s'_{t_2}, ..., s'_{t_k}\}_{R_i}$ and their corresponding competences obtained $\{\gamma_{s'_{t_1}}, \gamma_{s'_{t_2}}, ..., \gamma_{s'_{t_k}}\}_{R_i}$, indexed by their relative time order of experimentation $t_1 < t_2 < ... < t_k | t_{n+1} = t_n + 1$ inside this precise subspace R_i (t_i are not the absolute time, but integer indexes of relative order in the given subspace (region)).

An estimation of interest is computed for each region R_i. The interest $interest_i$ of a region R_i is described as *the absolute value of the derivative of local competences inside R_i, hence the local competence progress, over a sliding time window of the ζ more recent goals attempted inside R_i*:

$$CP(R_i) = \frac{\left(\sum_{j=|R_i|-\frac{\zeta}{2}}^{|R_i|-\frac{\zeta}{2}} \gamma_{s'_j} \right) - \left(\sum_{j=|R_i|-\zeta}^{|R_i|} \gamma_{s'_j} \right)}{\zeta}$$

And $interest_i = |CP(R_i)|$, where $CP(R_i)$ represents the Competence Progress inside R_i, using an absolute value to define $interest_i$ being used to consider cases of *increasing and decreasing competence*.

Indeed, an increasing competence signifies that the expected competence gain in R_i is important. We deduce that, potentially, selecting new goals in subspaces of high competence progress could bring on the one hand a high information gain for the learnt model, and on the other hand could lead to the reaching of not already reached goals. We call this phenomenon a positive intrinsic motivation.

Inversely, a decreasing competence in a region $\mathcal{R}_i$ means that some goals have been well reached the first time, but then the system has been less competent in reaching these same goals, or others situated in the same region. It can result from two different aspects of the considered region $\mathcal{R}_i$: first, different kinds of subregions are situated inside, some where goals can be accomplished, and others where the difficulty is too high according to the current learnt models; second, previously reached goals have become more difficult to achieve due to the release of constraints (described in the next sections). In opposition to the description of increasing competences, we define decreasing competences as providing negative motivation.

3) Goal Self-Generation Using the Measure of Interest: Using the previous description of interest, the goal self-generation and self-selection mechanism has to carry out two different processes:

1) Splitting of the space S' where goals are chosen into subspaces according to heuristics that allows maximal discrimination of areas according to their levels of interest;
2) Selecting the region where future goals will be chosen;

Such a mechanism has been described in the Robust-Intelligent Adaptive Curiosity (R-IAC) algorithm introduced in [10]. Here, we use the same kind of methods such as a recursive split of the space, each split being triggered once a maximal number of goals has been attempted inside. Each split is performed such that it maximizes the difference of the *interest* measure described above in the two resulting subspaces. This allows the easy separation of areas of different interests, and thus of different reaching difficulty.

Finally, goals are chosen according to the following heuristics which mixes three $modes$, once at least two regions exist after an initial random exploration of the whole space:

1. $mode(1)$: in $p_1\%$ percent (typically $p_1 = 70\%$) of goal selections. This algorithm chooses a random goal inside a region which is selected with a probability proportional to its interest value:

$$P_n = \frac{interest_n - \textbf{min}(interest_i)}{\sum_{i=1}^{|R_n|} interest_i - \textbf{min}(interest_i)}$$

Where P_n is the selection probability of the region R_n, and $interest_i$ corresponds to the current $interest$ of the region R_i.

2. $mode(2)$: in $p_2\%$ (typically $p_2 = 20\%$ of cases), this algorithm selects a random goal inside the whole space S'.

3. $mode(3)$: in $p_3\%$ (typically $p_3 = 10\%$), it first selects a region according to the interest value (like in $mode(1)$) and then generates a new goal close to the already experimented one which received the lowest competence estimation.

III. Biological Constraints and the Learning Process

Maturational constraints are considered as mechanisms evolving over the human life-time, allowing the control of the release of new capacities. While constraints are crucial for the evolution of infants, an important question about maturational constraints is how the maturational clock which determines the release and access of new capabilities evolves. While different studies about maturational constraints only considers preprogrammed continuous clocks [17], or discrete release depending on notions of learning saturation [26] or estimation of success [27], no system allowing a continuous active control of these constraints has been proposed in the literature.

In the following section, we present the Maturationally Constrained Self-Adaptive Goal Generation RIAC algorithm (McSAGG-RIAC^2) as a mechanism for controlling the evolution of maturational constraints by using heuristics that come from intrinsic motivations.

A. Maturational Constraints: the Role of Myelination

Maturational constraints play an important role in learning by partially determining a developmental pathway. Numerous biological reasons are part of this process such as the brain maturity, the weakness of infants' muscles or the development of the physiological sensory system. In McSAGG-RIAC^2 we take functional inspiration of constraints induced by brain maturation and especially by processes like **myelination** [28].

Related to the evolution of a substance called myelin, usually qualified by the term white matter, the main impact of myelination is to help the information transfer in the brain by increasing the speed at which impulses propagate along axons (connections between neurons). Here, we focus on the myelination process for several reasons. This phenomenon can indeed be considered as responsible for several maturational constraints: first it affects the motor development by limiting the frequency of feedback of electrical signals controlling

muscles; and second, it influences the processing of sensory signals received by the brain, such as sounds and visual images. By receiving an increasing amount of myelin over time, the infant thus access to a body and environment perceived with an increasing precision and thus complexity. In this study we consider only such increase in the physiological development, analysis of potential decrease which can arise over the development as proposed in [29] will be studied in future works.

In the following formalization, we consider constraints analogous to those induced by the myelination process, and propose a mechanism linking competence based intrinsic motivations with a maturational clock modeling the evolution of myelin responsible for the progressive release of constraints.

B. Formalization of Constraints

It is important to notice the multi-level aspect of maturational constraints: constraints existing on motor actions which influence the control, and by analogy in McSAGG-RIAC^2, the efficiency of the low-level active selection of actions performed to reach a goal; and constraints related to sensors like the capacity to perceive and discriminate objects, and thus here, to select a goal and/or declare it as reached.

Inspired by the increase of myelin appearing in the brain, we declare an evolving term $\psi(t)$ as a **maturational clock** responsible for the lifting of constraints. The main problem raised is to define a measure to control the evolution of this clock. For instance, in the Lift-Constraint, Act, Saturate (LCAS) algorithm [26], the authors use a simple discrete criteria based on a saturation threshold. They consider a robotic arm whose end-effector's position is observed in a task space. This task space is segmented into spherical regions of specified radius used as output for learning the forward kinematics of the robot. Each time the end-effector explores inside a spherical region, this region is activated. Once every region is activated, saturation occurs, and the radius of each region decreases so that the task space becomes segmented with a higher resolution and allows a more precise learning of the kinematics.

In the following section, we define a measure based on the competence progress which allows continuously and non-linearly controlling the maturational clock.

C. Stage Transition: Maturational Evolution and Intrinsic Motivations

The first version of McSAGG-RIAC introduced in [1] defines the evolution of the maturational clock as directly proportional to the global level of positive motivation. This leads the robot to potentially face with an overly complex environment, while sometimes releasing important quantities of constraints before the robot has managed to master the currently perceived environment. The measure of interest of SAGG-RIAC was also only defined using positive motivation (the absolute value defining the interest was not used), which prevents going back inside regions already visited, but where the release of constraints changed the level of complexity. In McSAGG-RIAC^2 we tackle these two issues and propose a robust definition of the release of constraints.

The main principle of bidirectional interactions between maturational constraints and learning in McSAGG-RIAC^2 is to increase $\psi(t)$ (lifting constraints) when the system is in a phase of decrease of its global intrinsic motivation, after having been positively motivated. Formally, it corresponds to periods of stabilization of the global competence level (estimation without consideration of regions), after a phase of progression (see Fig. 2). This stabilization is shown by a low derivative of the averaged competence level computed in the whole goal space S' in a recent time window $[t_{n-\frac{\zeta}{2}}, t_n]$ and the progression corresponds to an increase of these levels in a preceding time window $[t_{n-\zeta}, t_{n-\frac{\zeta}{2}}]$. Therefore, considering competence values estimated for the ζ last reaching attempts $\{\gamma_{s'_{n-\zeta}}, ..., \gamma_{s'_n}\}_{S'}$, $\psi(t)$ evolves until reaching a threshold ψ_{max} such that:

$$\psi(t+1) = \psi(t) + min\left(max_{evol}; \frac{\lambda}{CP(\{\gamma_{s'_{n-\zeta/2}}, ..., \gamma_{s'_n}\})}\right)$$

$$\text{if} \begin{cases} 0 < CP(\{\gamma_{s'_{n-\zeta/2}}, ..., \gamma_{s'_n}\}) < max_{CP} \\ CP(\{\gamma_{s'_{n-\zeta}}, ..., \gamma_{s'_{n-\zeta/2}}\}) > 0 \end{cases}$$

and $\psi(t+1) = \psi(t)$ otherwise, where max_{evol} is a threshold limiting a too rapid evolution of ψ, max_{CP} a threshold defining a stable competence level, and λ a positive factor. As the global evolution of the interest in the whole space is typically non-stationary, the maturational clock becomes typically non-linear and stops its progression when the global average of competence decreases. This decrease is due to the lifting of constraints, which increases the complexity of the perceived world.

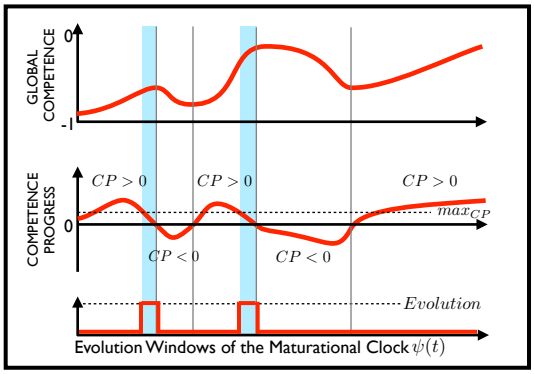

Fig. 2. Periods of evolution of the maturational clock $\psi(t)$ according to the evolution of the global competences.

D. Constraints Implementation

Maturational and morphological constraints can be described as genetically coded processes and can thus be defined as having emerged thanks to some evolutionary processes. Traditional mechanisms which carry out evolutionary processes use a fitness function which represent a precise goal that have to be optimized by a genetic/optimization algorithm. Here, we would like to consider robots that are able to learn high quantities of skills. Therefore, such a function cannot consider precise goals, but more the capability of the system to attain a maximal amount of goals.

In the following experiments, the evolution of constraints function of the maturational clock is handcrafted in order to show that bidirectional interactions between maturational constraints and learning allows the improvement of the efficiency of the learning and exploration processes.

IV. Experiment with a Quadruped Robot

A. Robotic Setup

In the following experiment, we consider a quadruped robot (see Fig. 3). Each of its legs is composed of 2 joints, the first

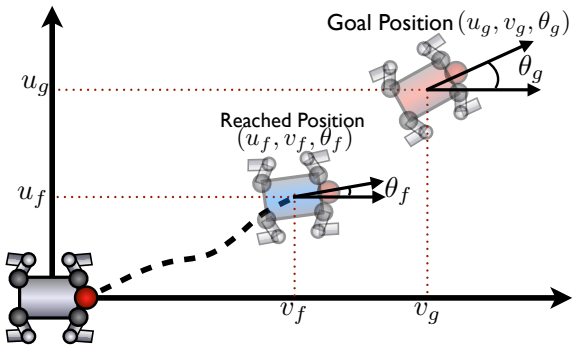

Fig. 3. Representation of the 12DOF quadruped, and measures used for the computation of competences: the goal and reached positions in (u, v, θ).

(closest to the robot's body) is controlled by two rotational DOF, and the second, one rotation (1 DOF). Each leg therefore consists of 3 DOF; the robot having in its totality 12 DOF. This robot is controlled using motor synergies Υ which directly specify the phase and amplitude of a sinusoid which controls the precise rotational value of each DOF over time [30]. These synergies are parameterized using a set of 24 continuous values, 12 representing the phase ph of each joint, and the 12 others, the amplitude am: $\Upsilon = \{ph_{1,2,..,12}; am_{1,2,..,12}\}$. Each experimentation consists of launching a motor synergy Υ for a fixed amount of time, starting from a fixed position. After this time period, the resulting position x_f of the robot is extracted into 3 dimensions: its position (u, v), and its rotation θ. The correspondence $\Upsilon \rightarrow (u, v, \theta)$ is then kept in memory as a learning exemplar.

The three dimensions u, v, θ are used to define the goal space of the robot and it is important to notice that precise areas reachable by the quadruped cannot be estimated beforehand. In the following experiment, we set the original dimensions of the goal space to $[-45; 45] \times [-45; 45] \times [-2\pi; 2\pi]$ on axis (u, v, θ), which was a priori larger than the reachable space. Then, after having carried out numerous experimentations, it appeared that this goal space was actually more than 25 times the size of the area accessible by the robot (see red contours in Fig. 4 where the reachable space is shown as inside $[-10; 10] \times [-10; 10] \times [-2\pi; 2\pi]$).

The implementation of our algorithm in such a robotic setup aims to test if the McSAGG-RIAC^2 driving method allows the robot to learn to attain a maximal amount of reachable positions, avoiding the selection of many goals inside regions which are unreachable, or that have previously been visited.

B. Measure of competence

In this experiment, we focus on the precise reaching of goal positions $x_g = (u_g, v_g, \theta_g)$. In every iteration the robot is reset to the same configuration called the origin position. We define the similarity function Sim and thus the competence as linked with the euclidian distance goal/robot $D(x_g, x_f)$ after a reaching attempt which is normalized by the original distance between the origin position x_{origin} and the goal $D(x_{origin}, x_g)$. This allows, for instance, the same competence level when considering a goal at 1km from the origin position which the robot approaches at 0.1km, and a goal at 100m which the robot approaches at 10m.

In this measure of competence, we consider the rotation factor θ, and compute the euclidian distance using (u, v, θ). Also, dimensions of the goal space are rescaled within [0;1]. Each dimension therefore has the same weight in the estimation of competence (an angle error of $\theta = \frac{1}{2\pi}$ is as important as an error $u = \frac{1}{90}$ or $v = \frac{1}{90}$). We formalize the similarity measure as the following: $Sim(x_g, x_f, x_{start}) = -\frac{D(x_g, x_f)}{D(x_{start}, x_g)}$ where $Sim(x_g, x_f, x_{start}) = 0$ if $D(x_{start}, x_g) = 0$.

C. Local Exploration and Reaching

Reaching a goal x_g necessitates the estimation of a motor synergy Υ_i leading to this chosen state x_g. Considering a single starting configuration for each experimentation, and motor synergies Υ, the forward model which defines this system can be written as $\Upsilon \rightarrow (u, v, \theta)$. Here, we have a direct relationship which only considers the 24 parameters $\{ph_{1,2,..,12}; am_{1,2,..,12}\}$ as inputs of the system, and a position in (u, v, θ) as output. Also, when considering the inverse model $(u, v, \theta) \rightarrow \Upsilon$ that has to be estimated, we use the following optimization mechanism which can be divided into two different phases: a reaching phase, and an exploration phase.

1) Reaching Phase: The reaching phase deals with reusing the data already learned to compute an inverse model $((u, v, \theta) \rightarrow \Upsilon)_L$ in the locality L of the intended goal $x_g = (u_g, v_g, \theta_g)$. In order to create such an inverse model (numerous can exist), we extract the potentially more reliable data using the following method: we first compute the set L of the l nearest neighbors of (u_g, v_g, θ_g) and their corresponding motor synergies using an ANN (Approximate Nearest Neighbors) method:
$L = \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_l\}$. Then, we consider the set M which contains l sets of m elements:
$M = \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_{1,2,..,l}$ where each set $\{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_i$ corresponds to the m nearest neighbors of each Υ_i, $i \in L$, and their corresponding resulting position (u, v, θ).

Finally, we select the set $O = \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}$ inside M such that it would be the one with the lowest standard deviation of its synergies in M. From O, we estimate a linear inverse model $((u, v, \theta) \rightarrow \Upsilon)$ by using a pseudo-inverse of Moore-Penrose, and obtain the synergy Υ_g which corresponds to the desired goal (u_g, v_g, θ_g).

2) Exploration Phase: The system here continuously estimates the distance between the goal x_g and the closest already reached position x_c. If the reaching phase does not manage to make the system come closer to x_g, i.e. $D(x_g, x_t) > D(x_g, x_c)$, with x_t as last experimented configuration in an attempt toward x_g, the exploration phase is triggered.

In this phase the system first considers the nearest neighbor $x_c = (u_c, v_c, \theta_c)$ of the goal (u_g, v_g, θ_g) and get the corresponding known synergy Υ_c. Then, it adds a random noise $rand(24)$ to the 24 parameters $\{ph_{1,2,..,12}, am_{1,2,..,12}\}_c$ of this synergy Υ_c which is proportional to the euclidian distance $D(x_g, x_c)$. The next synergy $\Upsilon_{t+1} = \{ph_{1,2,..,12}, am_{1,2,..,12}\}_{t+1}$ to experiment can thus be described as $\Upsilon_{t+1} = (\{ph_{1,2,..,12}, am_{1,2,..,12}\}_c + \lambda.rand(24).D(x_g, x_c))$ where $rand(i)$ is a vector of i random values in $[-1; 1]$, $\lambda > 0$ and $\{ph_{1,2,..,12}, am_{1,2,..,12}\}_c$ the motor synergy which corresponds to x_c.

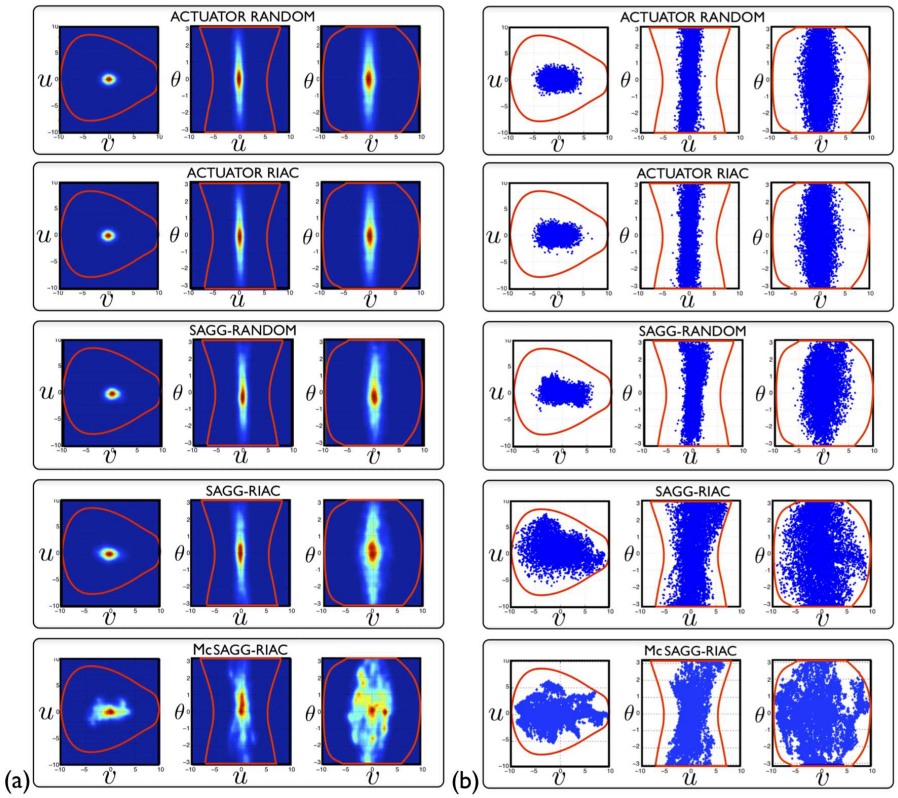

Fig. 4. Histograms of positions explored by the quadruped inside the goal space u, v, θ after 10000 experimentations (running a motor synergy during a fixed amount of time), using different exploration mechanisms.

D. Constraining the Goal Space

The goal space starts as a small sphere centered around the position $(u, v, \theta) = (0, 0, 0)$ which corresponds to the origin position where the quadruped starts each displacement. Then, according to the evolution of the maturational clock, the radius of this sphere increases, until it covers the entire goal space.

E. Constraining the Control Space

Due to the high number of parameters controlling each motor synergy, the learning mechanism faces a highly redundant system. Also, because our framework considers the fact of performing a maximal amount of tasks (i.e. goals) instead of different ways to perform a same task as important, constraints on the control space can be considered.

Let us consider the 24 dimensional space controlling phases and amplitudes as defined as $S = [-2\pi; 2\pi]^{12} \times [0; 1]^{12}$. We set the constrained subspace where possible values can be taken as $[\mu_i - 4\pi\sigma; \mu_i + 4\pi\sigma]^{12} \times [\mu_j - \sigma; \mu_j + \sigma]^{12} \in S$, where μ corresponds to a seed, different for each dimension, around which values can be taken according to a window of size 2σ; σ varying according to the maturational clock $\psi(t)$.

We aim to show the potential increase of efficiency of the learning process that can arise thanks to the coupling of constraints in the control and goal spaces. To reveal this increase, we do not need to optimize the value of each seed according to complex mechanism. In this experiment, we handcrafted the value of each seed according to the simple following rule: first, we run an experiment only using constraints in the goal space. Once this experiment terminated, we compute histograms of phases and amplitude experimented with during the exploration process. Then, the seed selected for each dimension corresponds to the maximum of the histogram, which represents the majority value used during this experiment.

F. Qualitative Results

In [13], we presented only qualitative results while using an early version called McSAGG-RIAC with a robotic arm. We showed the self-adaptive behavior of the algorithm, which is able to actively accelerate and slow down the evolution of the maturational clock when considering constraints whose release evolves with different velocities. In the following section, we propose qualitative and quantitative analysis with a new robotic setup, whose limits of reachability are impossible to predict in advance, contrary to the experiment presented in [13]. For each experiment, we verify that the release of constraints is developed enough in order to cover the whole reachable area at the end of the experiment. This permits performing rigorous comparisons with methods which do not use constraints.

Fig. 4 (a), presents representative examples of histograms of positions explored by the quadruped inside the goal space u, v, θ after 10000 experimentations (running of motor synergies during the same fixed amount of time), and (b) shows examples of the repartitions of positions inside the goal space after 10000 experimentations when using the following exploration mechanisms:

ACTUATOR-RANDOM corresponds to a uniform selection of parameters controlling motor synergies (values inside the 24 dimensional space of phases and amplitudes). ACTUATOR-RIAC corresponds to the original version of the R-IAC algorithm presented in [10] which actively generates actions inside the same space of synergies as ACTUATOR-RANDOM. SAGG-RANDOM is a method where the learning is situated at

the level of goals which are generated uniformly in the goal space u, v, θ. Here the low-level of active learning used is the same as in SAGG-RIAC. Then, the SAGG-RIAC method corresponds to the self-generation of goals actively inside the whole goal space while McSAGG-RIAC^2 also considers maturational constraints in both control and goal spaces.

Comparing the two first exploration mechanisms (ACTUATOR-RANDOM and ACTUATOR-RIAC) we cannot distinguish any notable difference, the space explored appears similar and the extent of explored space on the (u, v) axis is comprised in the interval $[-5; 5]$ for u and $[-2.5; 2.5]$ for v on both graphs. Nevertheless, these results are important when comparing histograms of exploration (Fig. 4 (a)) and visited positions (Fig. 4 (b)) to the size of the reachable area (red lines on Fig. 4). It indeed shows that, in the 24 dimensional space controlling motor synergies, an extremely large part of values lead to positions close to $(0, 0, 0)$, and thus do not allow the robot to perform a large displacement. It allows the deduction that reaching the entire goal space is a difficult task which could be discovered using exploration in the space of motor synergies only after extremely long time periods. Moreover, we notice that the difference between u and v scales is due to the inherent structure of the robot which simplifies the way to go forward and backward rather than shifting left or right.

Considering SAGG methods, it is important to note the difference between the reachable area and the goal space. In Fig. 4, red lines correspond to the estimated reachable area which is comprised of $[-10; 10] \times [-10; 10] \times [-\pi; \pi]$, whereas the goal space is much larger: $[-45; 45] \times [-45; 45] \times [-2\pi; 2\pi]$. We are also able to notice the asymmetric aspect of its repartition according to the v axis, which is due to the decentered weight of the robot's head.

The SAGG-RANDOM method seems to slightly increase the space covered on the u and v axis compared to ACTUATOR methods, as shown by the higher concentration of positions explored in the interval $[-5; -3] \cup [3; 5]$ of u. However, this change does not seem very important when comparing SAGG-RANDOM to any previous algorithm.

SAGG-RIAC, contrary to SAGG-RANDOM, shows a large exploration range compared to other methods: the surface in u has almost twice as much coverage than using previous algorithms, and in v, up to three times; there is a maximum of 7.5 in v where the previous algorithms were at 2.5. These last results emphasize the capability of SAGG-RIAC to drive the learning process inside reachable areas which are not easily accessible (hardly discovered by chance). Nevertheless, when observing histograms of SAGG-RIAC, we can notice the high concentration of explored positions around $(0, 0, 0)$, the starting position where every experimentation is launched. This signifies that, even if SAGG-RIAC is able to explore a large volume of the reachable space, as shown in Fig. 4 (b), it still spends many iterations exploring the same areas.

According to the repartition of positions shown in Fig. 4 (b) for the McSAGG-RIAC^2 exploration mechanism, we can first notice a volume explored comparable to the one explored by SAGG-RIAC. Nevertheless, it seems that McSAGG-RIAC^2 visits a slightly lower part of the space, avoiding some areas, while explored area seems to be visited with a higher concentration. This higher concentration is confirmed via observation of histograms of McSAGG-RIAC^2; indeed, whereas every other method focuses a large part of their exploration time around the position $(0, 0, 0)$, McSAGG-RIAC^2 also focuses in areas distant from this position. The higher attraction toward different areas is due to the fixation of constraints in the goal space: limiting the goal space allows a fast focalization of the algorithm toward reachable goals, whereas without constraints, the system spends large amounts of time attempting unreachable goals before discriminating reachable areas, and thus performs movements which have a high probability of leading to positions close to $(0, 0, 0)$ according to the kinematics of the system. Also, areas visited a few times such as the upper right part of the third graph of McSAGG-RIAC^2 Fig. 4 (b), can be explained by the high focalization of McSAGG-RIAC^2 in other areas, as well as the constraints limiting the values taken in the 24 dimensional control space.

G. Quantitative Results

In this section, we aim to test the efficiency of the learned database to guide the quadruped to reach a set of goal positions. Here we consider a test database of 100 goals and compute the distance between each goal attempted, and the reached position. Fig. 5 shows performances of methods introduced previously. Also, in addition to the evaluation of the efficiency of McSAGG-RIAC^2 with constraints in both control and goal spaces (called McSAGG-RIAC^2 In&Out in Fig. 5), we introduce the evaluation of McSAGG-RIAC^2 when only using constraints on the goal space (McSAGG-RIAC^2 Out).

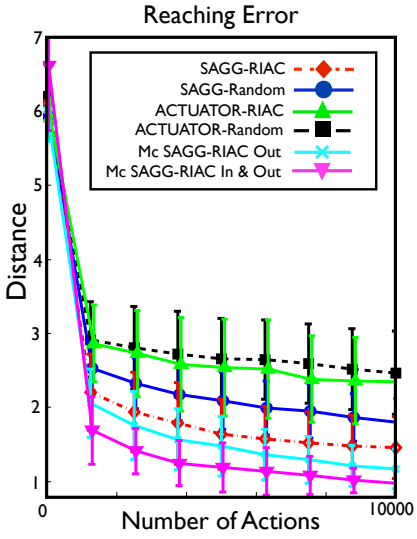

Fig. 5. Reaching errors estimated using different databases collected with different exploration methods on a over 10 quadruped experiments of 10000 actions.

First of all, we can observe the higher efficiency of SAGG-RIAC compared to the other three methods which can be observed after only 1000 iterations. The high decreasing velocity of the reaching error (in the number of experimentations) is due to the consideration of regions limited to a small number of elements (30 in this experiment). It allows the creation of a very high number of regions within a small interval of time, which helps the system to discover and focus on reachable regions and its surrounding area.

ACTUATOR-RIAC shows slightly more efficient performances than ACTUATOR-RANDOM. Also, even if SAGG-RANDOM is less efficient than SAGG-RIAC, we can observe its highly decreasing reaching errors compared to ACTUATOR methods, which allows it to be significantly more efficient than these method when considered at 10000 iterations.

McSAGG-RIAC^2 Out shows better results than SAGG-RIAC since the beginning of the evolution (1000 iterations), and decreases with a higher velocity until the end of the experiment. This is due to the fast discovery of reachable areas (showed in Fig. 4 (a)) which reduces the exploration of synergies leading to positions close to $(0, 0, 0)$, and leads to a more uniform exploration inside the reachable space. This emphasizes first the high potential of coupling constraints situated in the goal space and SAGG-RIAC in such a complex robotic setup.

Eventually, we can observe that coupling constraints in both control and goal spaces as introduced by McSAGG-RIAC^2 In & Out, obtains significantly more efficient results than SAGG-RIAC without constraints ($p = 0.0055$ at the end of the exploration process), and better than when only using constraints in the goal space with a measure of significance of $p = 0.05$. Constraining the 24 dimensional space which controls motor synergies thus allows an important simplification of the learning process, while allowing the efficient reaching of a number of goals as important as without constraints.

In such a highly-redundant robot, coupling different types of constraints with the SAGG-RIAC process thus obtains significantly better performances than when using the SAGG-RIAC competence based intrinsic motivation algorithm without it. This one being significantly more efficient than the other methods proposed, including the original RIAC algorithm.

These experiments emphasize the high efficiency of methods which drive the exploration at the level of goals. SAGG methods, and especially SAGG-RIAC, permit driving the exploration in order to explore large spaces containing areas hardly discovered by chance, when limits of reachability are impossible to predict. Then, thanks to the fixation of constraints in the goal space, McSAGG-RIAC^2 manages to direct the exploration more uniformly than SAGG-RIAC.

Eventually, quantitative results showed the capability of SAGG-RANDOM and SAGG-RIAC methods to learn inverse models efficiently when considering highly-redundant robotic systems controlled with motor synergies. Then, the high focalization of McSAGG-RIAC^2 in areas visited a few by SAGG-RIAC results in a improved learning efficiency.

V. Conclusion

In this paper we argued that intrinsic motivations and maturational constraints mechanisms might have complex bidirectional interactions which actively control the growth of complexity in motor development. We proposed an integrated system of these two frameworks which allows a robot to developmentally learn its inverse model progressively and efficiently, and presented qualitative and quantitative results showing the high potential of the McSAGG-RIAC^2 algorithm to direct an efficient learning process when considered as an active learning algorithm.

VI. Acknowledgment

This research was partially funded by ERC Grant EXPLORERS 240007.

References

[1] A. Baranes and P.-Y. Oudeyer, "Maturationaly-constrained competence-based intrinsically motivated learning," in *Proceeding of the IEEE International Conference on Development and Learning (ICDL)*, 2010.

[2] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 54 – 67, 2000.

[3] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11(2), pp. pp. 265–286, 2007.

[4] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[5] V. Fedorov, *Theory of Optimal Experiment*. New York, NY: Academic Press, Inc., 1972.

[6] S. Thrun, "Exploration in active learning," in *Handbook of Brain Science and Neural Networks*, M. Arbib, Ed. Cambridge, MA: MIT Press, 1995.

[7] K. Merrick and M. L. Maher, "Motivated learning from interesting events: Adaptive, multitask learning agents for complex environments," *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*, vol. 17, no. 1, pp. 7–27, 2009.

[8] J. Schmidhuber, "Curious model-building control systems," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 2, 1991, pp. 1458–1463.

[9] J. Schmidhuber, "Adaptive curiosity and adaptive confidence," Institut fur Informatik, Technische Universitat Munchen, Tech. Rep. FLI-149-91, 1991.

[10] A. Baranes and P.-Y. Oudeyer, "Riac: Robust intrinsically motivated exploration and active learning," *IEEE Transation on Autonomous Mental Development*, vol. 1, no. 3, pp. 155–169, 2009.

[11] M. Schembri, M. Mirolli, and B. G., "Evolution and learning in an intrinsically motivated reinforcement learning robot," in *Advances in Artificial Life. Proceedings of the 9th European Conference on Artificial Life*, Springer, Ed., Berlin, 2007, pp. 294–333.

[12] J. Schmidhuber, "Optimal artificial curiosity, developmental robotics, creativity, music, and the fine arts," *Connection Science*, vol. 18, no. 2, 2006.

[13] A. Baranes and P. Y. Oudeyer, "Intrinsically motivated goal exploration for active motor learning in robots: A case study," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, 2010.

[14] M. Schlesinger, "Heterochrony: It's (all) about time!" in *Proceedings of the Eighth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, L. U. C. Studies, Ed., Sweden, 2008, pp. 111–117.

[15] E. Thelen, D. M. Fisher, and R. Ridley-Johnson, "The relationship between physical growth and a newborn reflex," *Infant Behavior and Development*, vol. 7, pp. 479–493, 1984.

[16] N. Bernstein, *The Coordination and Regulation of Movements*. Pergamon, 1967.

[17] Y. Nagai, M. Asada, and K. Hosoda, "Learning for joint attention helped by functional development," *Advanced Robotics*, vol. 20, no. 10, pp. 1165–1181, September 2006.

[18] R. Pfeifer and C. Scheier, *Understanding Intelligence*. Cambridge, MA: MIT Press, 1999.

[19] P. J. Bentley and S. Kumar, "Three ways to grow designs: A comparison of embryogenies for an evolutionary design problem." in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999)*, 1999, pp. 35–43.

[20] J. C. Bongard, "Evolving modular genetic regulatory networks," in *Proceedings IEEE 2002 Congress on Evolutionary Computation (Inst of Electrical Engineers, London),*, 2002, pp. 1872–1877.

[21] M. Lungarella and L. Berthouze, "Adaptivity through physical immaturity," in *Proc. of the 2nd Int. Workshop on Epigenetic Robotics*, 2002.

[22] P. Oudeyer and F. Kaplan, "How can we define intrinsic motivations ?" in *Proc. Of the 8th Conf. On Epigenetic Robotics.*, 2008.

[23] M. Rolf, J. Steil, and M. Gienger, "Goal babbling permits direct learning of inverse kinematics," *IEEE Trans. Autonomous Mental Development*, vol. 2, no. 3, pp. 216–229, 09/2010 2010.

[24] J. Peters and S. Schaal, "Natural actor critic," *Neurocomputing*, no. 7-9, pp. 1180–1190, 2008. [Online]. Available: http://www-clmc.usc.edu/publications//P/peters-NC2008.pdf

[25] S. Nolfi and D. Floreano, *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press, 2000.

[26] M. Lee, Q. Meng, and F. Chao, "Staged competence learning in developmental robotics," *Adaptive Behavior*, vol. 15, no. 3, pp. 241–255, 2007.

[27] J. C. Bongard, "Morphological change in machines accelerates the evolution of robust behavior," *Proceedigns of the National Academy of Sciences of the United States of America (PNAS)*, January 2010.

[28] J. Eyre, *Development and Plasticity of the Corticospinal System in Man*. Hindawi Publishing Corporation, 2003.

[29] M. Lungarella and L. Berthouze, "Adaptivity via alternate freeing and freezing of degrees of freedom," in *Proc. of the 9th Intl. Conf. on Neural Information Processing*, 2002.

[30] A. Ijspeert, "Central pattern generators for locomotion control in animals and robots: A review," *Neural Networks*, vol. 21, no. 4, pp. 642–653, 2008.