The Mathematical Theory of Embodied Agentic AI

Daniel Drake

February 8, 2026

1 Symbols

Let the following hold:

- $\mathbb{N}_K = \{1, 2, 3, ..., K\}$
- $\mathbb{N}_K^0 = \{0\} \cup \mathbb{N}_K$
- $T \subseteq \mathbb{R}$
- $card(T) = N \in \mathbb{N}$
- $T = \{t_0, t_1, ..., t_f\}$ where $(\forall j \in \mathbb{N}_{N-1}^0)(t_j < t_{j+1})$
- T represent the times at which different events happen.
- $T_K = \{t_0, t_1, ..., t_K\} \subseteq T$ represents the first K events.

1.1 Sensors

Let the following hold:

- $X = \prod_{j \in \mathbb{N}_K} X_j$ such that $(\forall j \in \mathbb{N}_K)(X_j \text{ is a TVS})$
- $(\forall x \in X)(x = (x_1, ..., x_K))$
- $\pi_j : X \to X_j \pitchfork \pi_j(x) = x_j$
- $mem(type(x)) = \sum_{j \in \mathbb{N}_K} mem(type(\pi_j(x)))$
- S represent a sensor.
- $S : T \to X$
- $D = S(T)$ represents all the data collected by the sensor.
- $D_k = S(T_k)$ represents the first k captures of the sensor.

Table 1: 16 Fundamental operations

A	0	1	0	1
B	0	0	1	1
False	0	0	0	0
A AND B	0	0	0	1
$A \not\Leftarrow B$	0	0	1	0
B	0	0	1	1
$A \not\Rightarrow B$	0	1	0	0
A	0	1	0	1
A XOR B	0	1	1	0
A OR B	0	1	1	1
A NOR B	1	0	0	0
A Equivalent B	1	0	0	1
Not A	1	0	1	0
$A \Rightarrow B$	1	0	1	1
Not B	1	1	0	0
$A \Leftarrow B$	1	1	0	1
A NAND B	1	1	1	0
True	1	1	1	1

1.1.1 Proposition

Given a finite amount of information, we can only draw a finite number of possible conclusions from it. If we had two variables each of which can store 1 bit of information. Then there are 16 possible manipulations of these variables and no more. It takes 4 bits of information to represent the operator associated with any of these manipulations. We are left with a problem: Since we can know all of these operators: How do we figure out which operators are useful?

In robotics we have this problem. We have sensor input and we need to transform that sensor input into packets of usable information. We often have high dimensional data with is to complicated to deal with on it's own. From this we need to find ways to effectively go from the high dimensional data to a low dimensional data which is easier to deal with.

In addition to this we have another problem. A single frame of high dimensional data might not be enough information to appropriatly deal with that frame of data. We might need previous frames to come to the correct conclusion about the current frame.

A **Goal** is a function on which to optimize.
A **Strategy** is a solution to that optimization problem.
An **Implementation** is the collection of details on which the strategy is implemented.
A **Context** is the collection of details around a {Goal, Strategy, Implementation} problem.

Example:

- $W, H \in \mathbb{N}$
- $I = [0,1]^{WxH}$
- $(x, y, z \in I)(x \neq z)$
- $f_z : I \to \mathbb{R}$
- $(\forall y \in I)(f_z(y) = ||y - z||)$
- $M \in \mathbb{N}$
- Goal: Find any path: $\{x_t\}_{t \in \mathbb{N}_M^0}$ such that: as t goes from $0 \to M$, x_t goes from $x \to z$
- Strategy:
  $x_t = (\frac{t}{M} * z) + ((1 - \frac{t}{M}) * x)$
  $f_z(x_0) = ||(\frac{0}{M} * z) + (1 - \frac{0}{M}) * x - z|| = ||x - z||$
  $f_z(x_M) = ||(\frac{M}{M} * z) + (1 - \frac{M}{M}) * x|| = ||z - z|| = ||0|| = 0$
- Implementation:

  `size_t` t ▷ Loop counter
  for $t = 0$ **to** M **do**
  $x_t = \frac{t}{M} z + \left(1 - \frac{t}{M}\right) x$
  end for

Goals::

- Find an operator T which turns a sequence of images x_t into a the 6 dof localization of the camera. Roll, Pitch, Yaw, X, Y, Z.
- Find an operator T which creates value in an economic system.
- Find an operator T which turns a sequence of images x_t into a 3d point cloud with identifications of the objects in the point clouds.
- Find an operator T which turns a sequence of images x_t into a collection of sentences generated from words written in the scene.
- Find an operator T which passively listens for a wakeup word.
- Find an operator T which actively listens for spoken word for a period of time.
- Find an operator T which can turn text into text.
- Find an operator T which can turn text into sounds.
- Find an operator T which can turn the general neural network activity into sounds.
- Find an operator T which can generate new GSI problems involving 3d point clouds.
- Find an operator T which can create the semblance of moods.
- Find an operator T which tries to engage in natural conversation.
- Find an operator T which tries to teach the droid language.
- Find an operator T which can coordinate and direct which goals get executed.
- Find an operator T which tries to translate verbal instruction into a GSI problem.
- Find an operator T which executes artifical general intelligence.
- Find an operator T which executes free will.
- Find an operator T which documents a GSI problems into UML.
- Find an operator T which allows the robot to explore and map a space.
- Find an operator T which allows the robot to climb on surfaces.
- Find an operator T which will allow the robot to follow another object.
- Find an operator T which filters a GSI problem through Issac Asimov's laws of robotics.
- Find an operator T which tune the robot's servos and walking motion through a sensor position solution.
- Find an operator T which calculates the forward kinematics.
- Find an operator T which calculates the inverse kinematics.
- Find an operator T which does laser reflection + camera information analysis.

- Find an operator T which does a power usage analysis.
- Find an operator T which recognizes animal behaviors.
- Find an operator T which allows the robot to dig.
- Find an operator T which identifies different materials.
- Find an operator T which can actively try to solve GSI problems.

2 Time of flight image based localization