Learning Iterative Reasoning through Energy Minimization
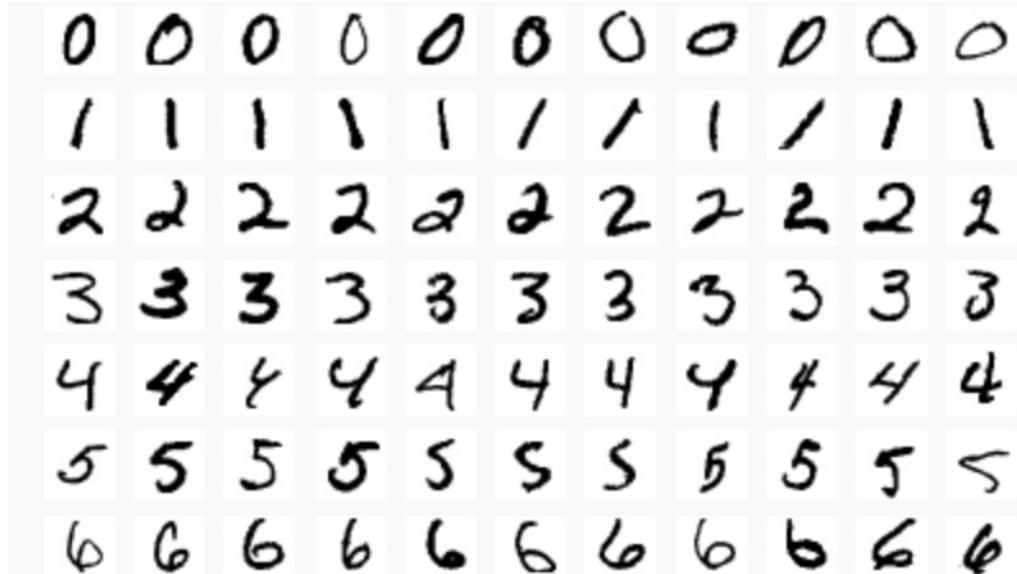
Yilun Du[1], Shuang Li[1], Joshua B. Tenenbaum[1,2], Igor Mordatch[3]
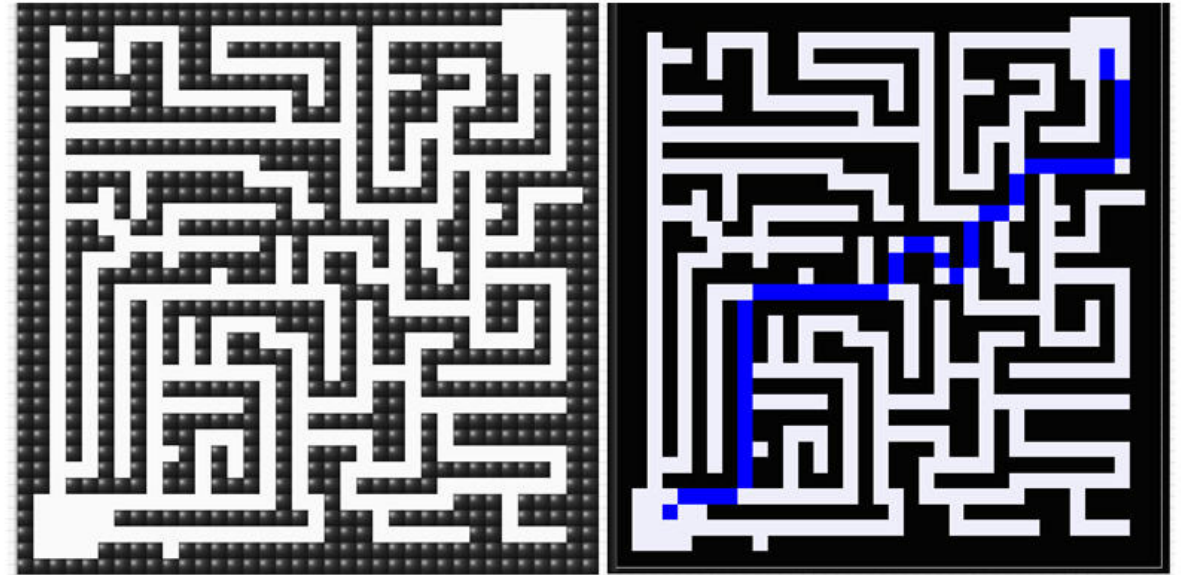
[1]MIT CSAIL, [2]MIT BCS, [3]Google Brain

Goal: Flexible Reasoning in the World
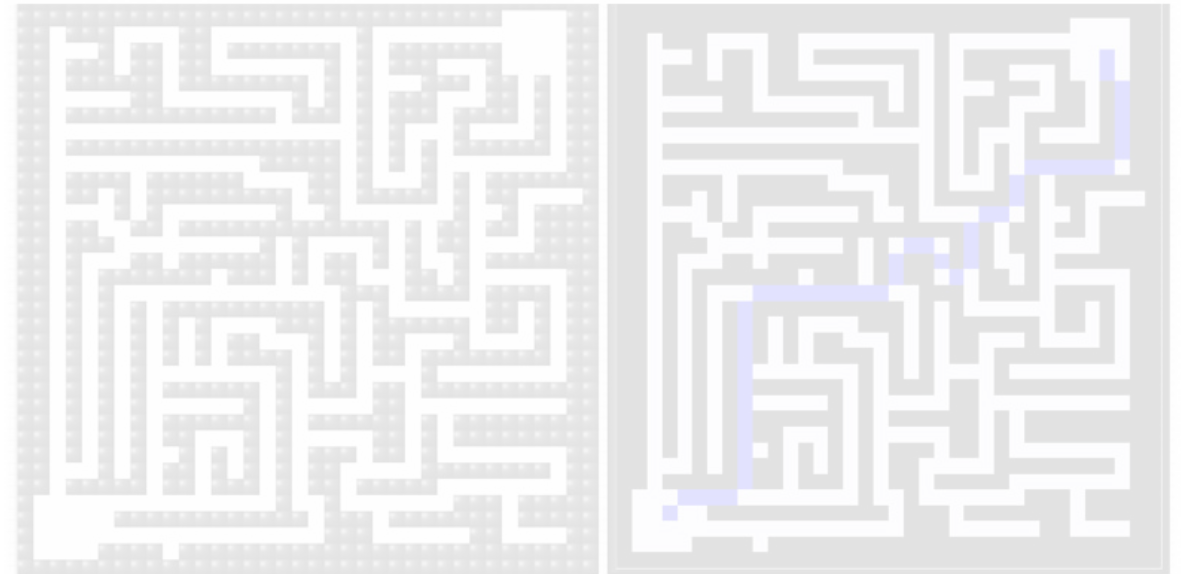

System I
(Pattern Recognition)

System II
(Procedural Reasoning)
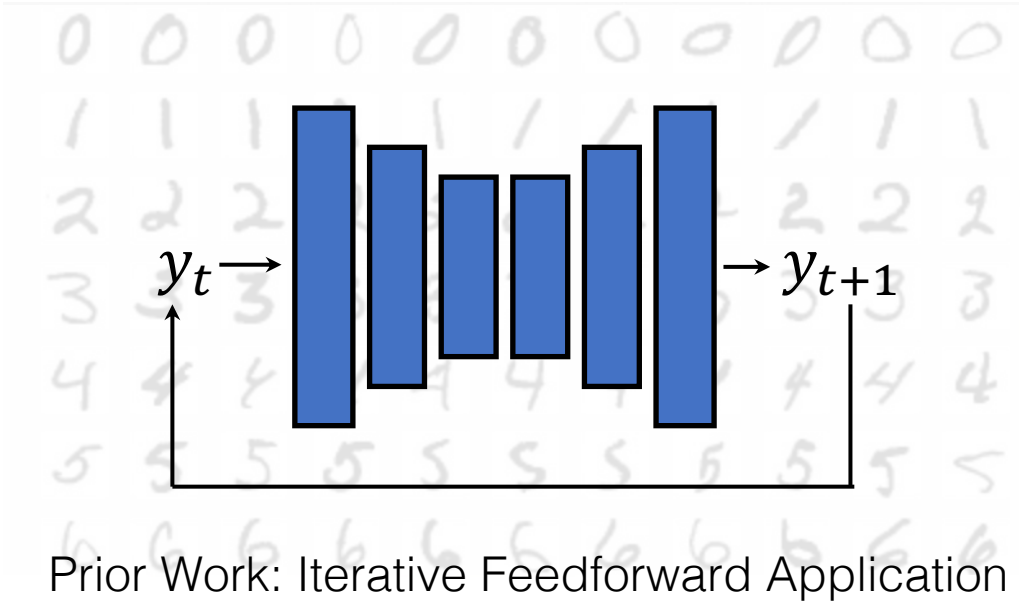
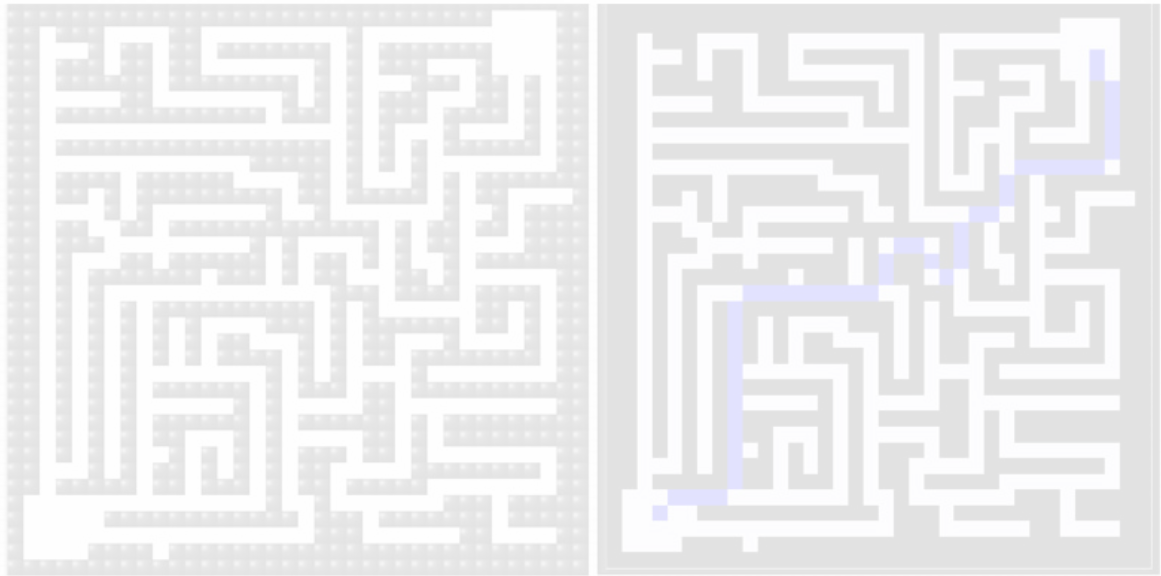# Goal: Flexible Reasoning in the World

System I
(Pattern Recognition)

System II
(Methodological Reasoning)

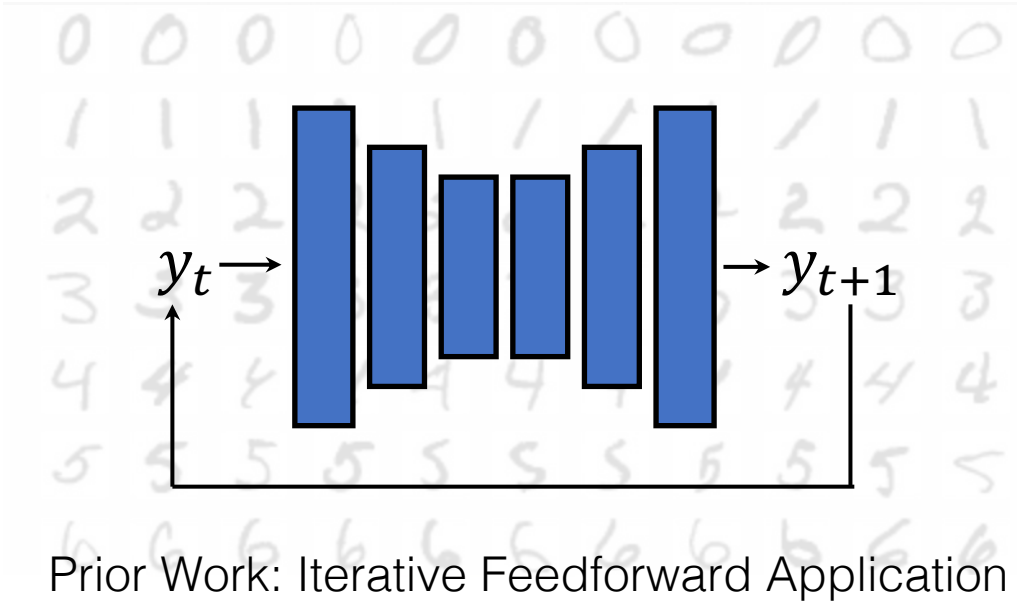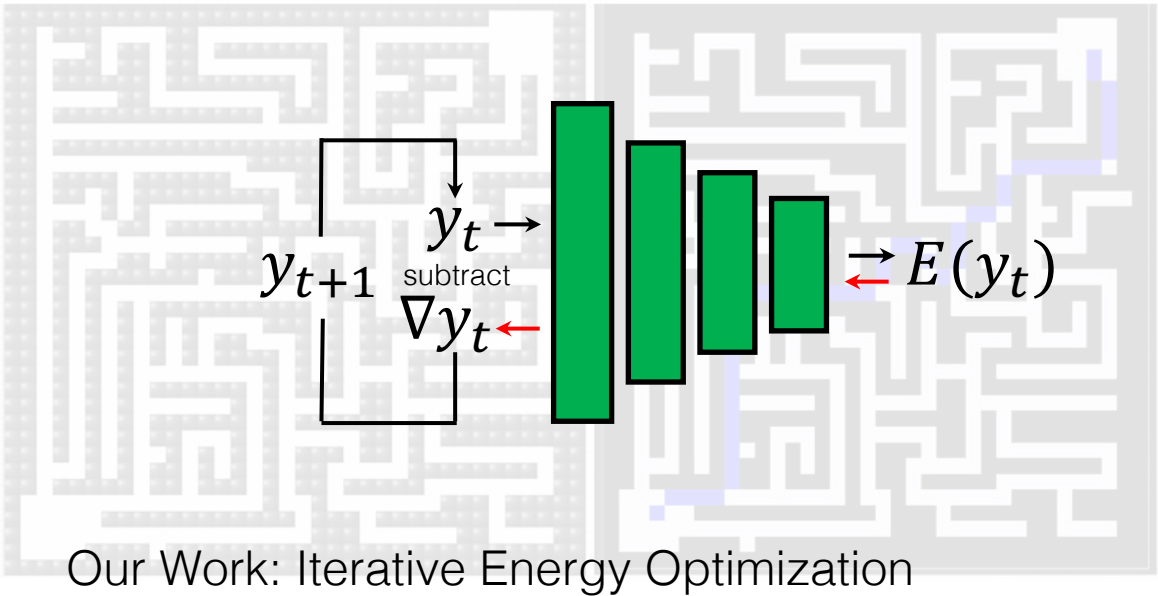# Goal: Flexible Reasoning in the World

Prior Work: Iterative Feedforward Application

System I
(Pattern Recognition)

System II
(Methodological Reasoning)

Goal: Flexible Reasoning in the World

Prior Work: Iterative Feedforward Application
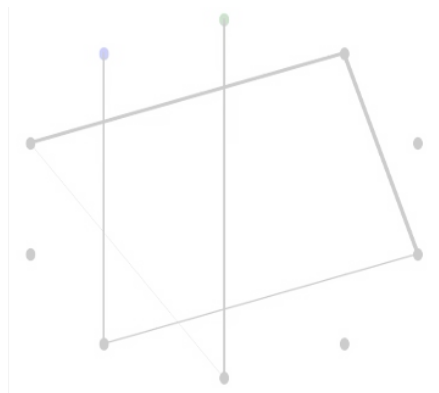
System I
(Pattern Recognition)

Our Work: Iterative Energy Optimization

System II
(Methodological Reasoning)

Reasoning as Energy Minimization
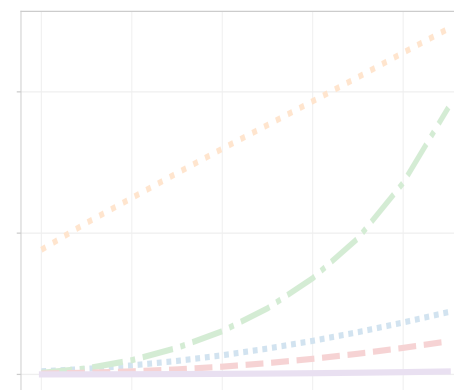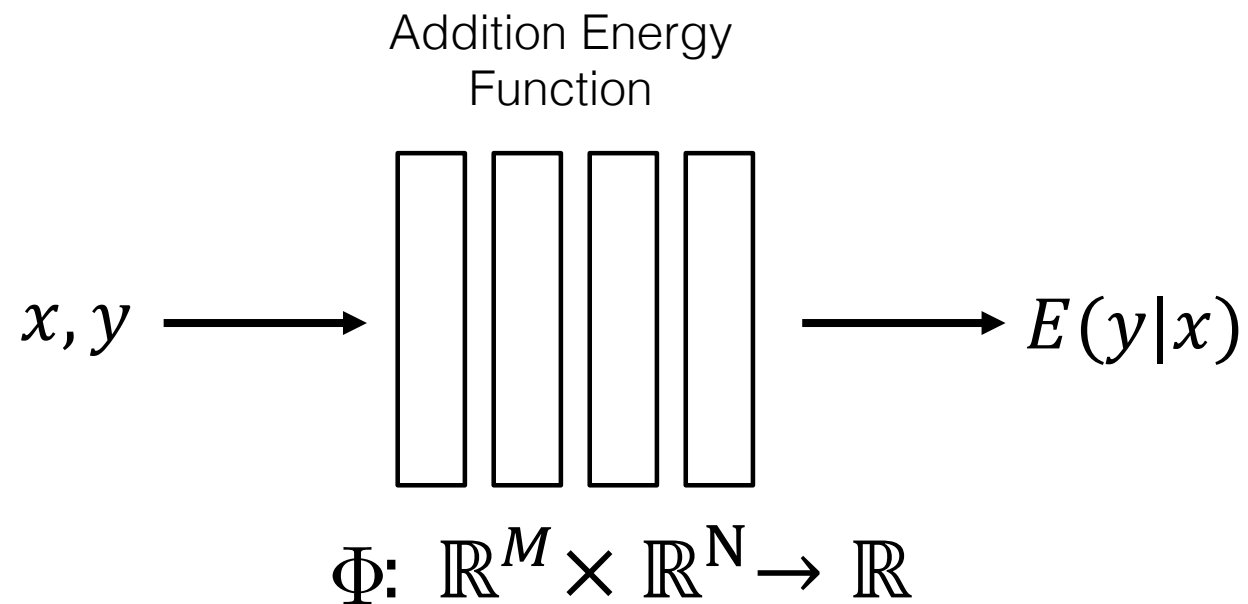
$E_\theta(z|x,y)$
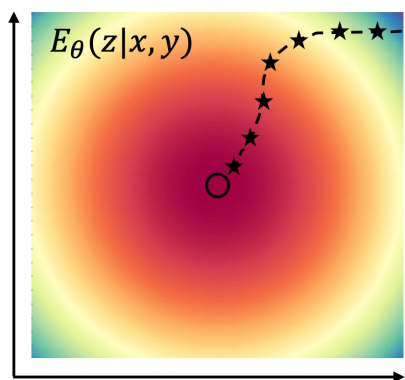
Graph Reasoning

Matrix Reasoning

Recursive Reasoning

Reasoning through Energy Minimization
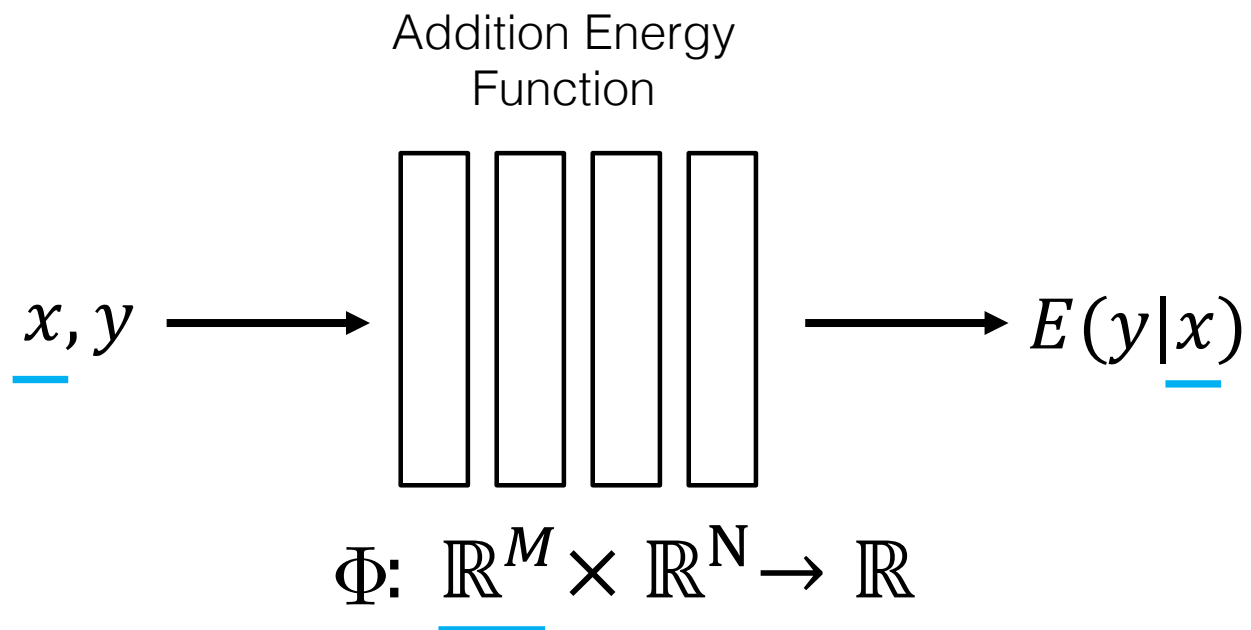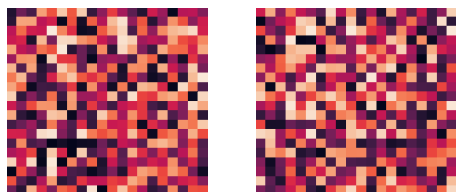
Addition Energy
Function

$$x, y \longrightarrow \quad E(y|x)$$

$$\Phi: \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$$

Reasoning through Energy Minimization

Inputs:

Addition Energy
Function

x, y ⟶ [ ] [ ] [ ] [ ] ⟶ $E(y|x)$

$\Phi: \ \mathbb{R}^M \times \mathbb{R}^N \rightarrow \mathbb{R}$
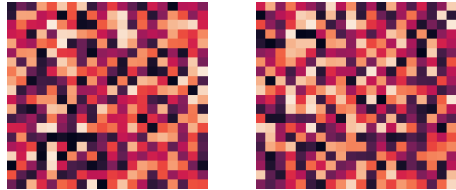
Reasoning through Energy Minimization

Inputs:

Solutions:

Addition Energy Function

$$x, y \longrightarrow \boxed{} \longrightarrow E(y|x)$$

$$\Phi: \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$$
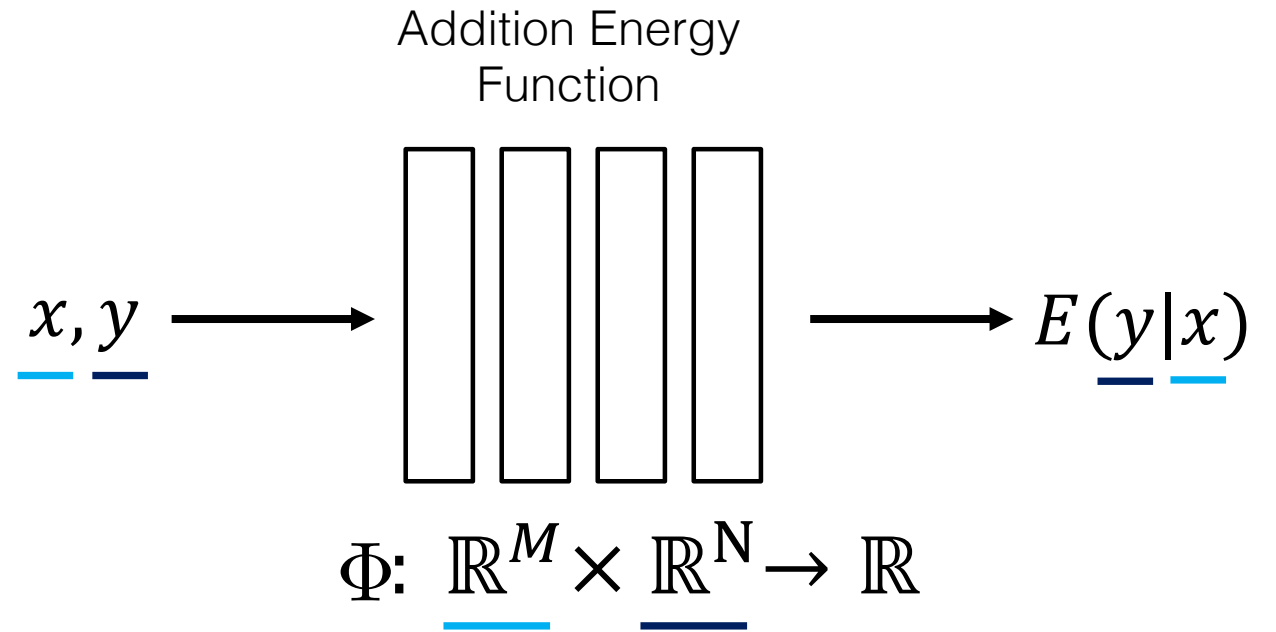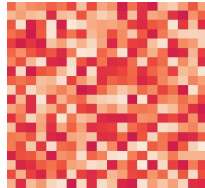
Reasoning through Energy Minimization

Inputs:

Solutions:

  Incorrect Answer

Addition Energy Function

$x, y \longrightarrow$

$\longrightarrow$ High Energy

$E(y|x)$

$\Phi: \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$
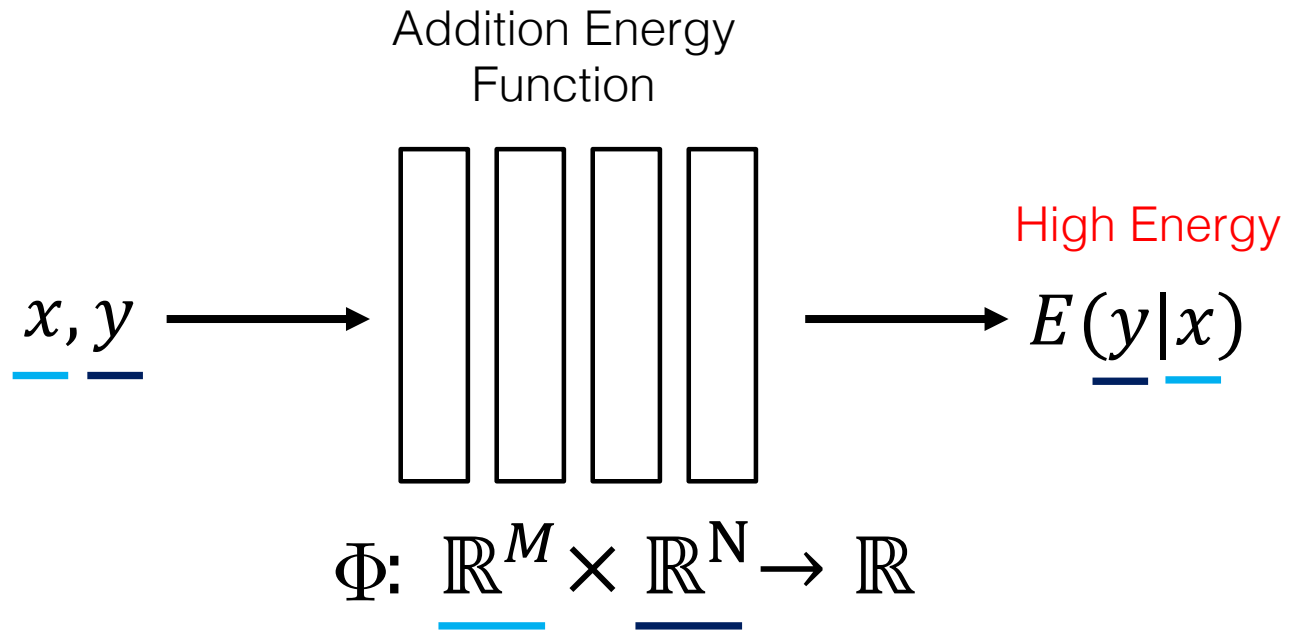
Reasoning through Energy Minimization

Inputs:

Solutions:

Correct Answer
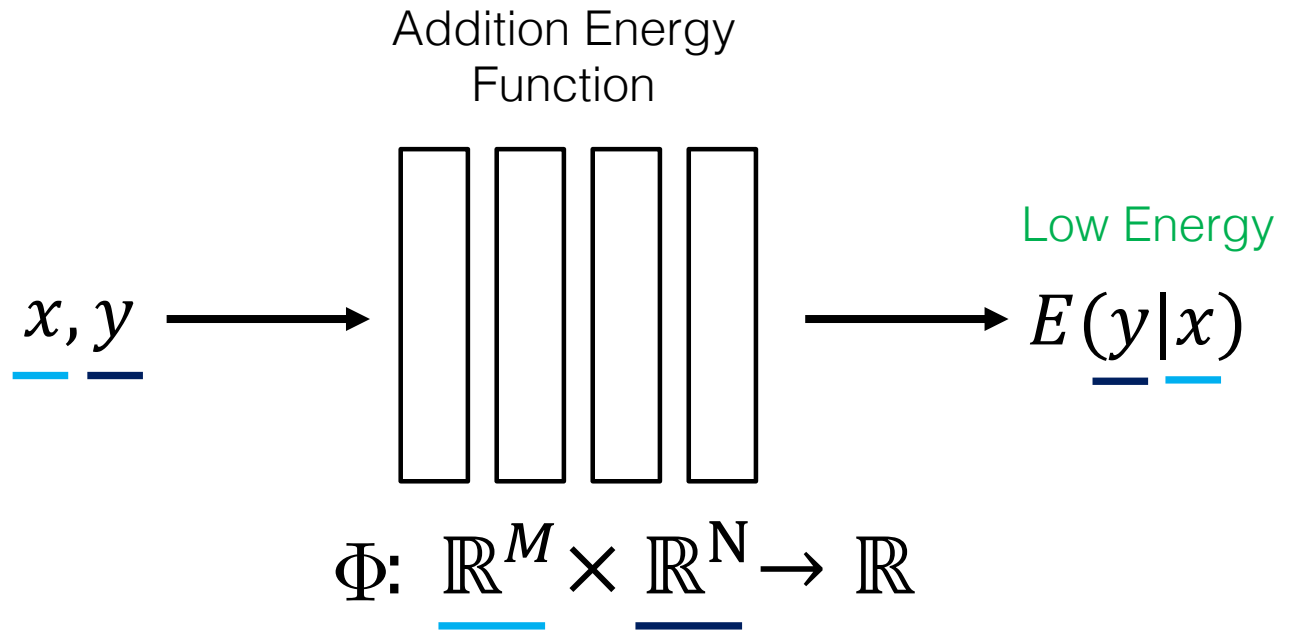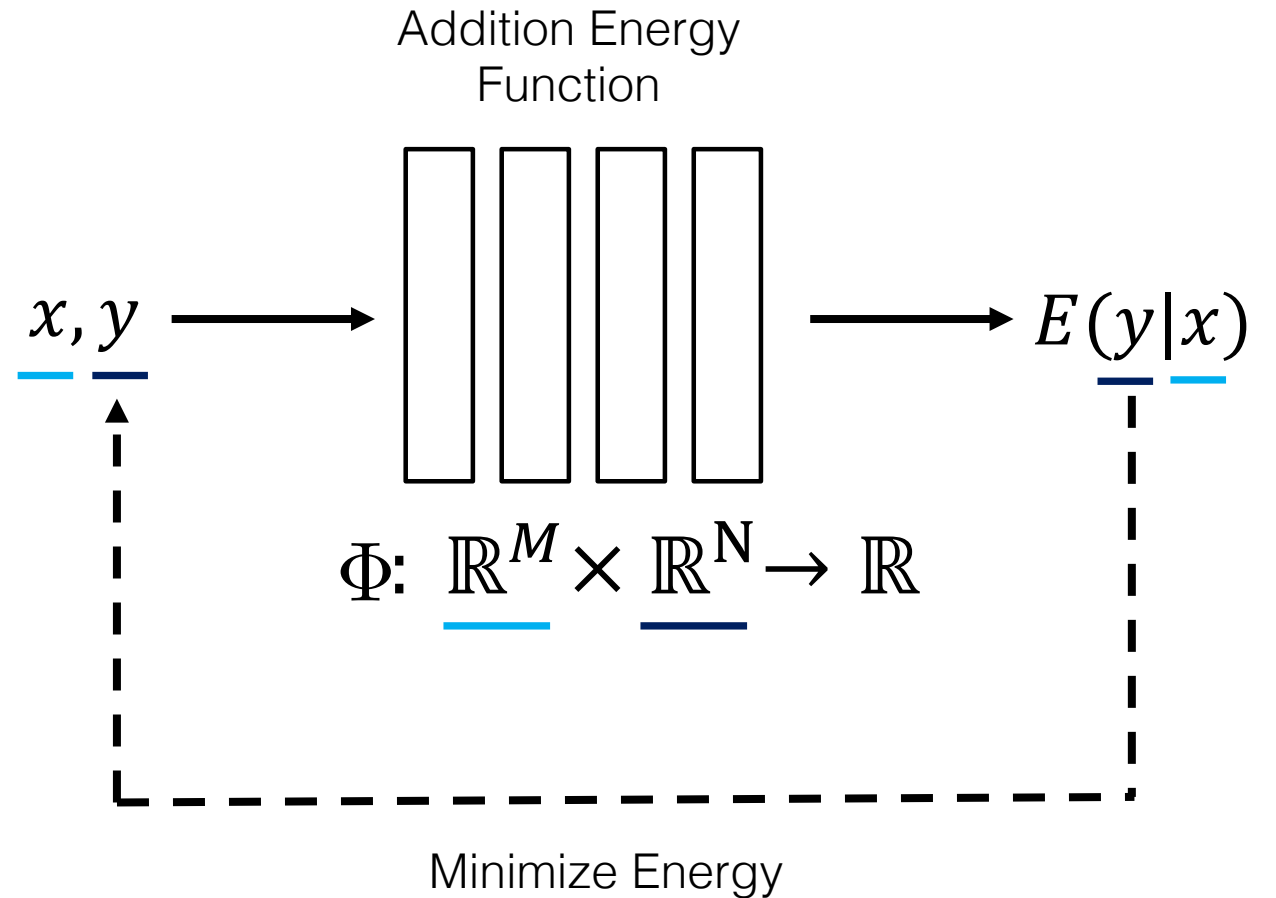
Addition Energy Function

$x, y \longrightarrow$

Low Energy

$E(y|x)$

$\Phi: \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$

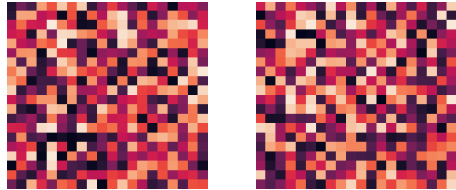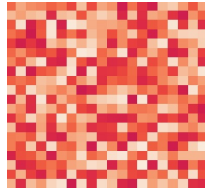# Reasoning through Energy Minimization
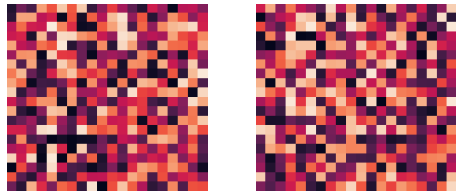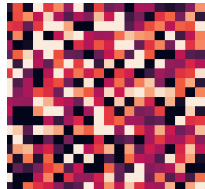
Inputs:

Solutions:

Addition Energy
Function

$$x, y \longrightarrow \quad \longrightarrow E(y|x)$$

$$\Phi: \ \mathbb{R}^M \times \mathbb{R}^N \rightarrow \mathbb{R}$$

Minimize Energy
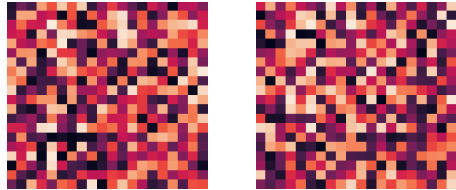
Reasoning through Energy Minimization

Inputs:

Solutions:

Addition Energy
Function

$$x, y \longrightarrow \boxed{| | | |} \longrightarrow E(y|x)$$

$$\Phi: \ \mathbb{R}^M \times \mathbb{R}^N \rightarrow \mathbb{R}$$

Minimize Energy

Reasoning through Energy Minimization

Inputs:

Solutions:

Supervise Optimization

Addition Energy Function

$$x, y \longrightarrow \boxed{} \longrightarrow E(y|x)$$

$$\Phi: \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$$
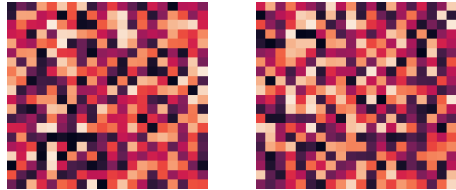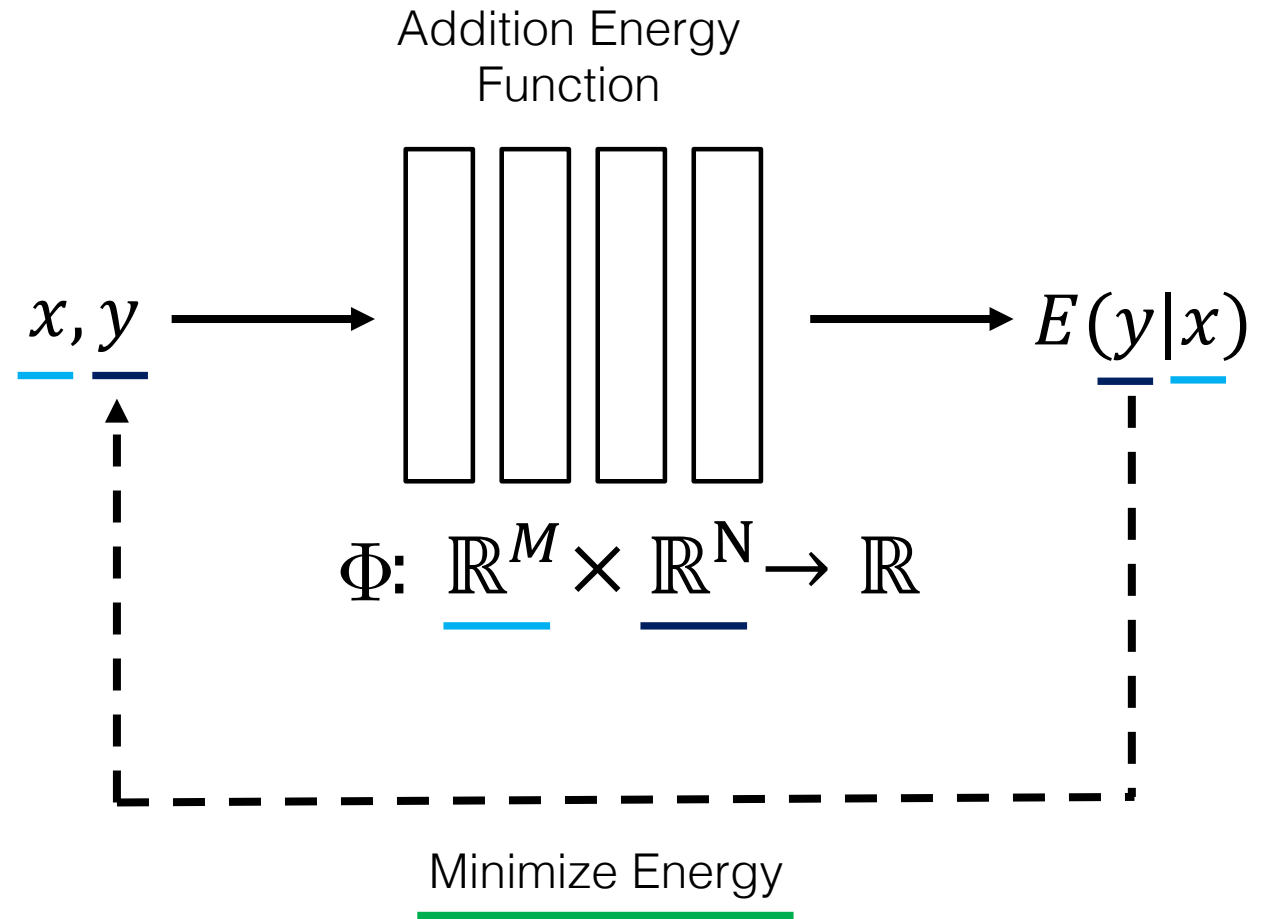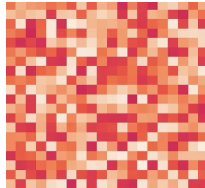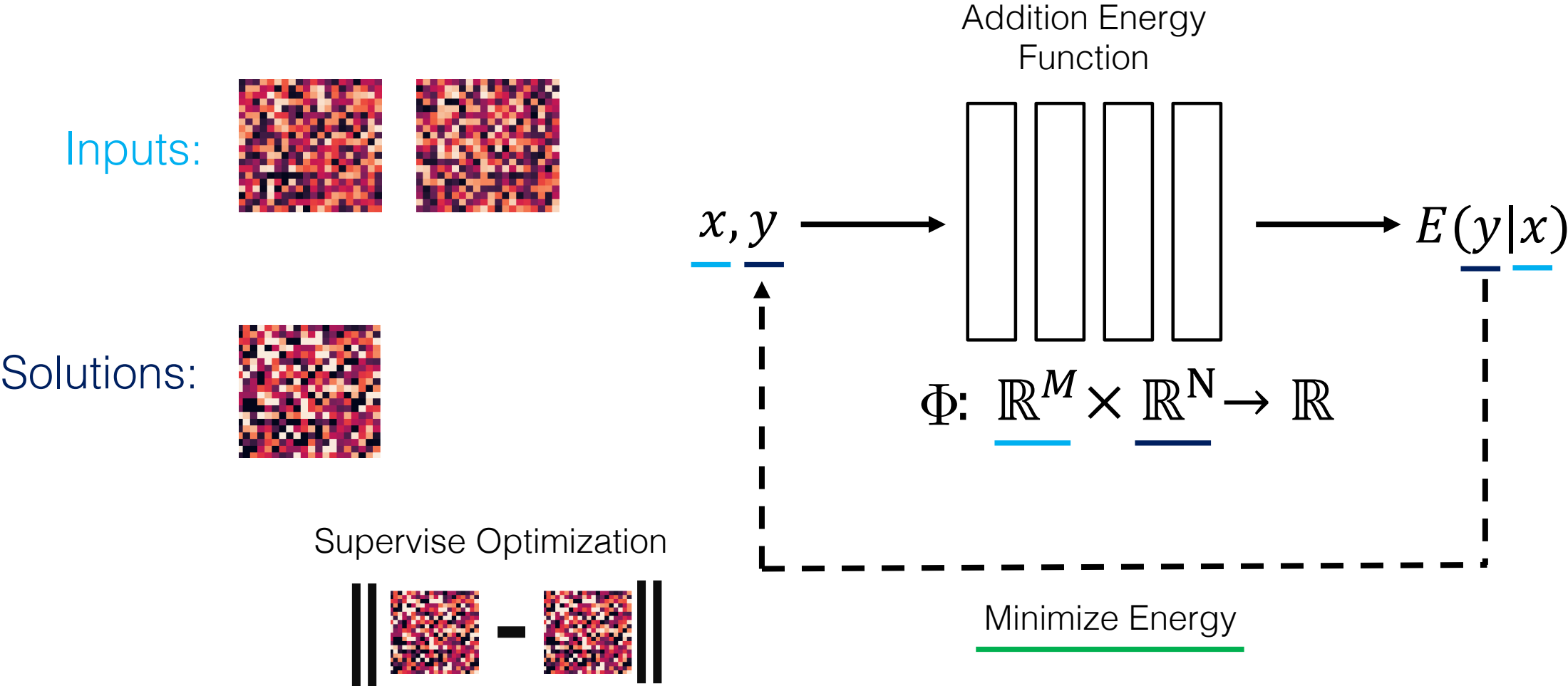
Minimize Energy

Reasoning through Energy Minimization

Summation Algorithm

Sum(x, y)

Input: input $x \in \mathbb{R}^K$
        input $y \in \mathbb{R}^K$
Output: output $z \in \mathbb{R}^K$
Initialization: $z \leftarrow [0]^K$
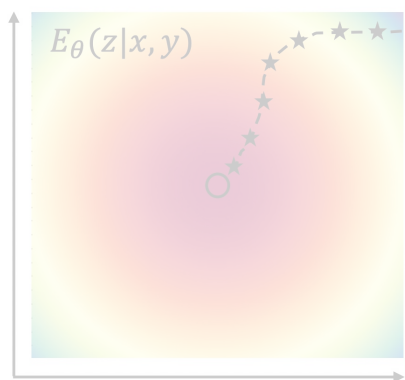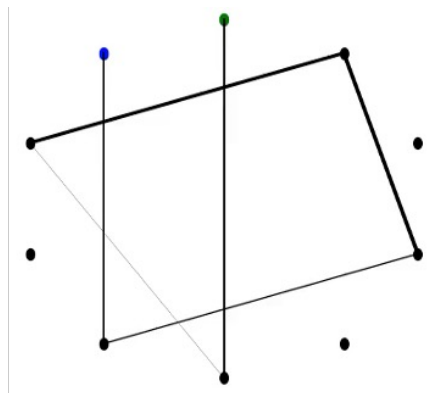**for** index $k = 1$ to K **do**
    $z_k \leftarrow x_k + y_k$
end for

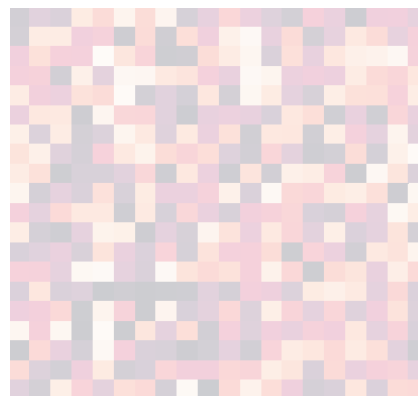$$\mathbf{Sum}(\boldsymbol{x}, \boldsymbol{y}) = \arg\min_{\boldsymbol{z}} E_\theta(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$$
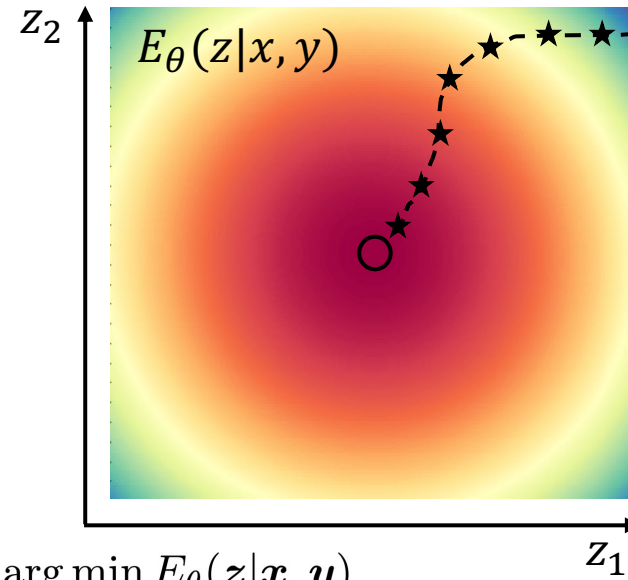
Reasoning as Energy
Minimization
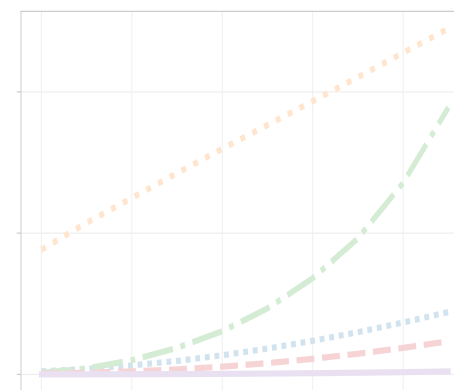
Graph
Reasoning

Matrix
Reasoning

Recursive
Reasoning

$E_\theta(z|x,y)$

$E_\theta(z|x,y)$

Graphical Reasoning
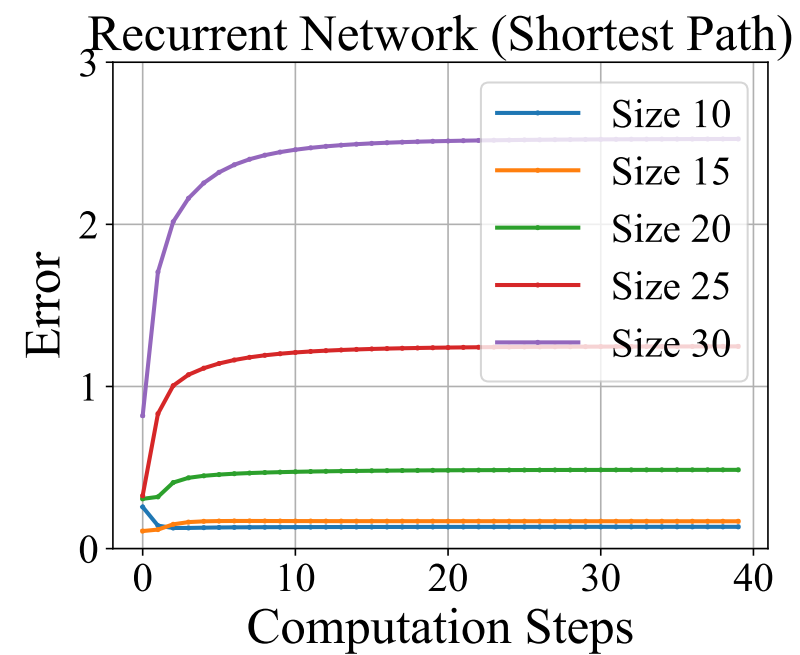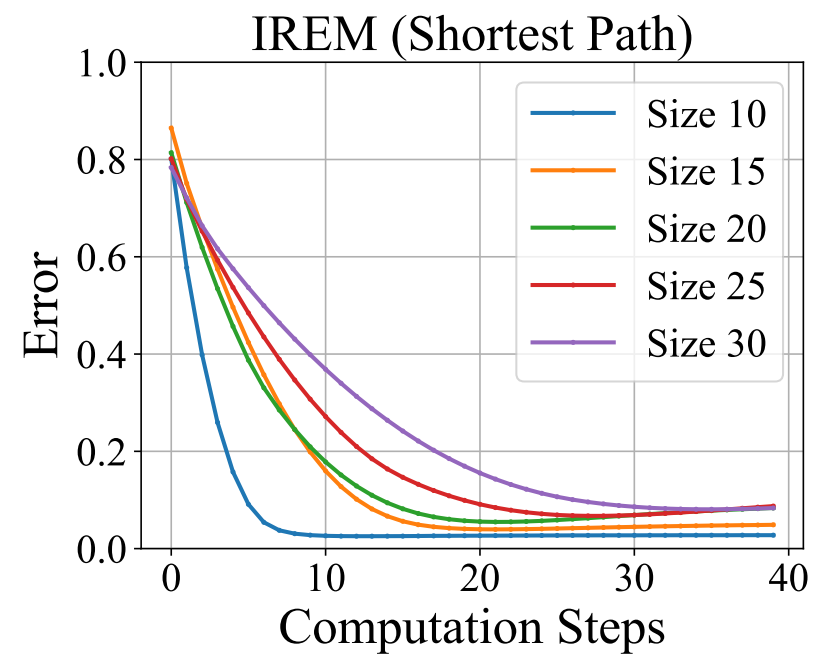
Graphical Reasoning

Similar trends in connected components
and edge copy.

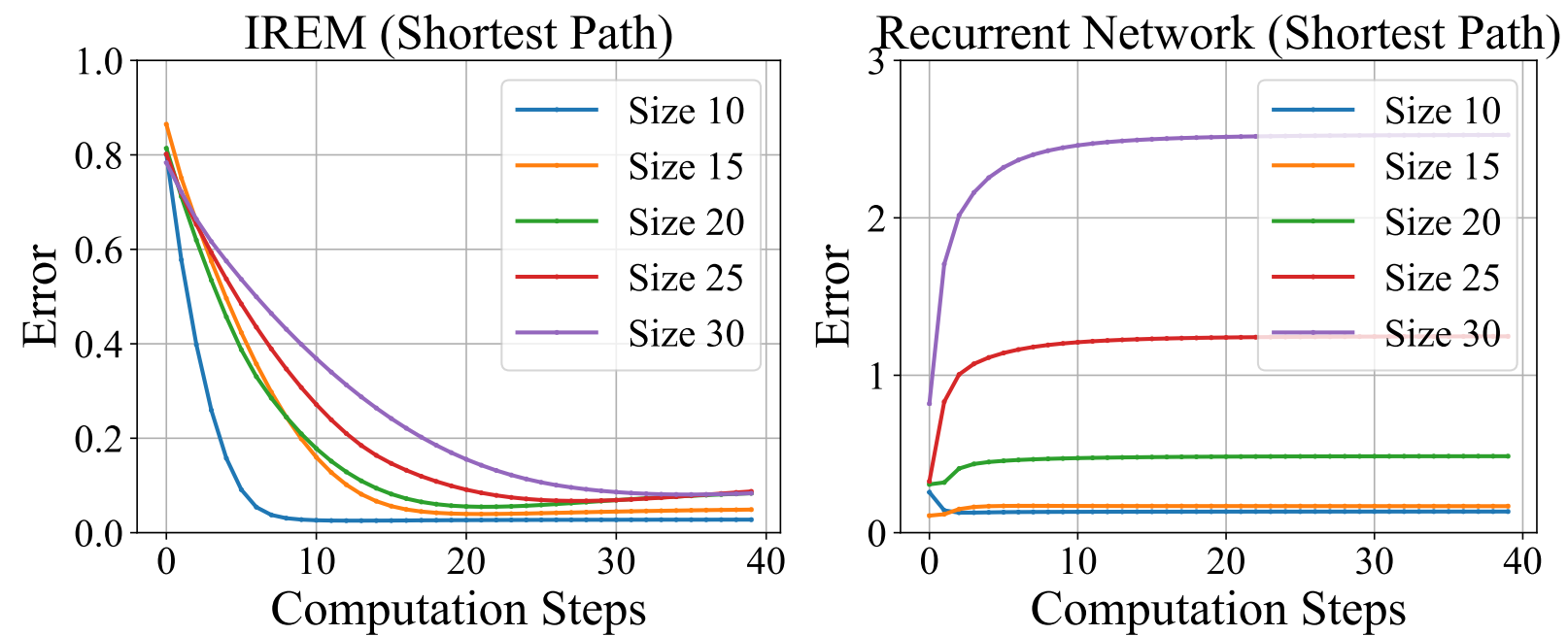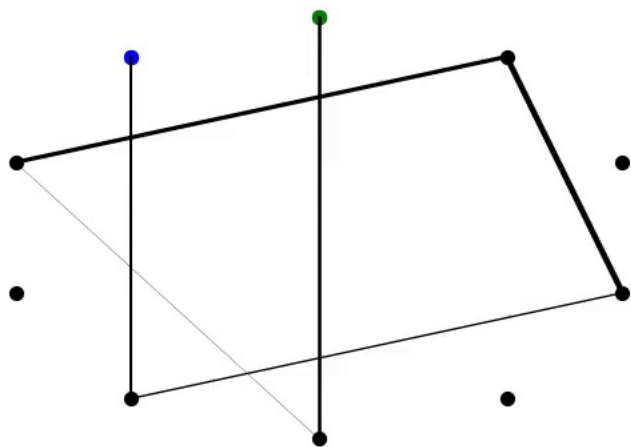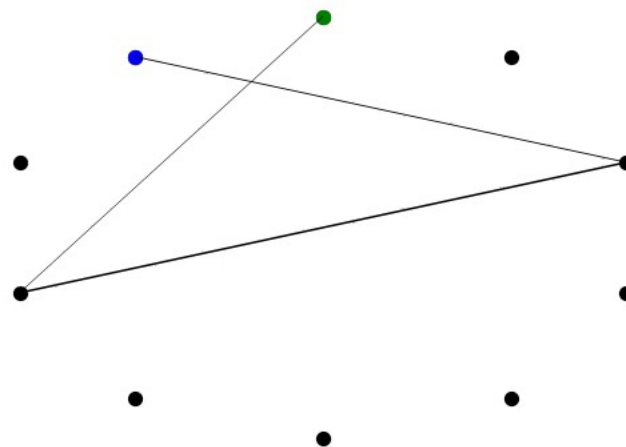# Graphical Reasoning

Predicted Shortest Path (Distance 3.338)   True Shortest Path (Distance 0.685)

Graphical Reasoning

Predicted Shortest Path (Distance 2.439)

True Shortest Path (Distance 0.685)

Graphical Reasoning
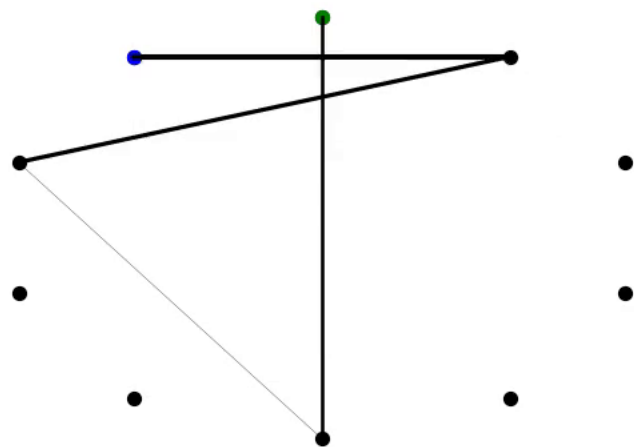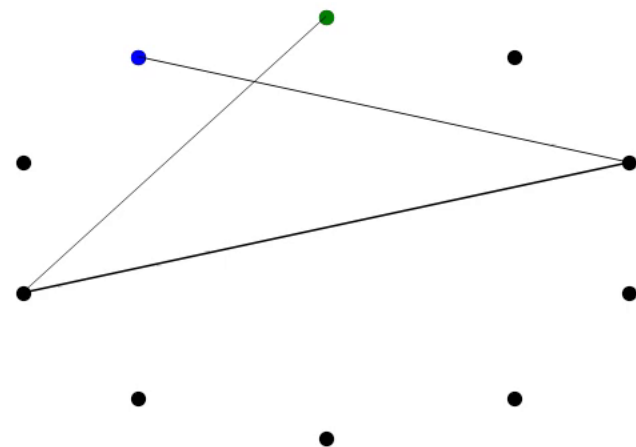

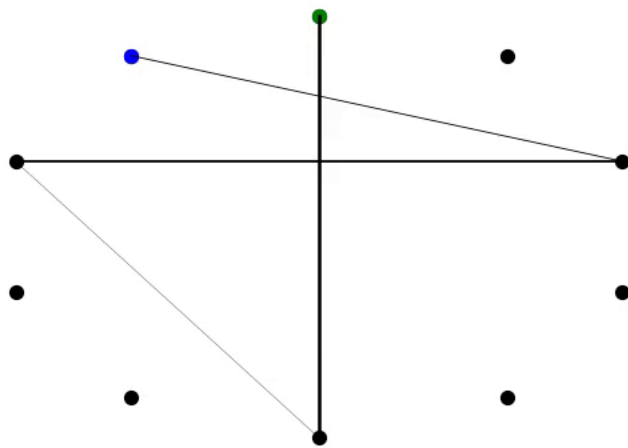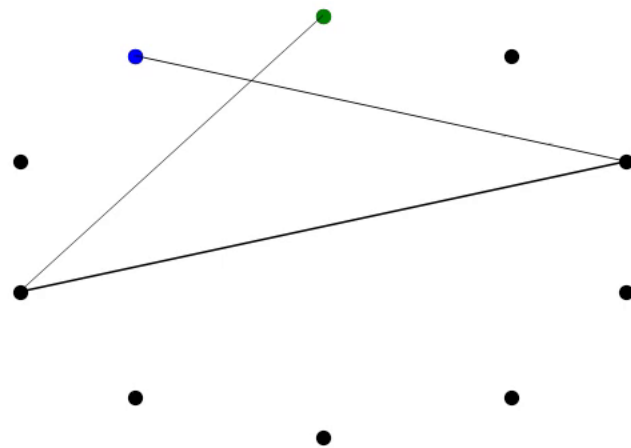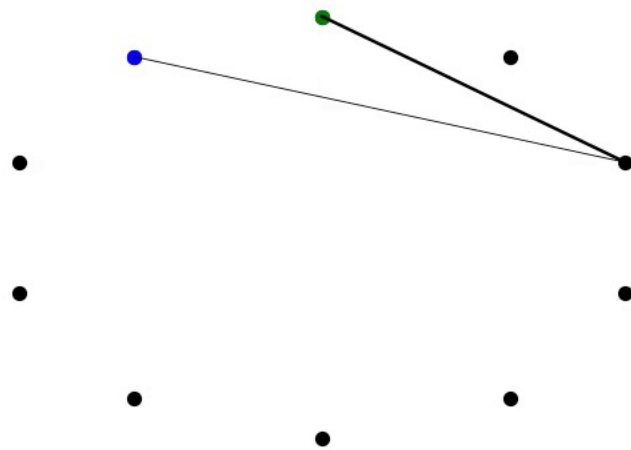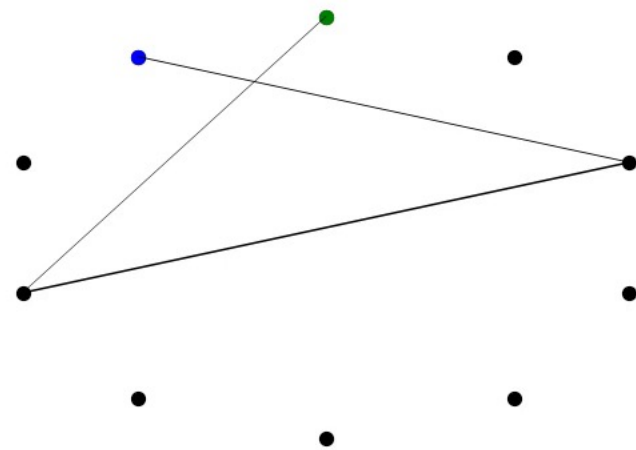Predicted Shortest Path (Distance 1.330) True Shortest Path (Distance 0.685)

Graphical Reasoning

Predicted Shortest Path (Distance 0.790)   True Shortest Path (Distance 0.685)

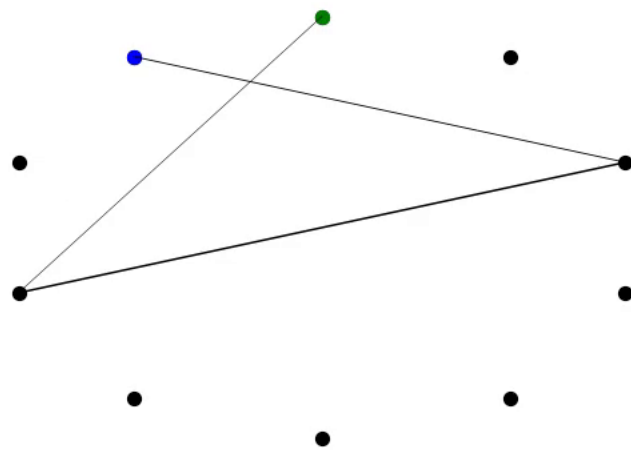# Graphical Reasoning

Predicted Shortest Path (Distance 0.685)   True Shortest Path (Distance 0.685)

Reasoning as Energy Minimization — $E_\theta(z|x,y)$

Graph Reasoning

Matrix Reasoning

Recursive Reasoning

Error
Error
Error (Log Scale)
Error
Computation Steps
Computation Steps
Computation Steps
Computation Steps

Iterative Feedforward   IREM (Ours)
IREM (Ours)

Matrix Completion
Error
Computation Steps

Matrix Inverse
Error
Computation Steps

Matrix Completion
Error
Comput

Matrix Completion (Harder)
Matrix Inverse
Error
Computation Steps

Matrix Inverse (Harder)
Error
Computation Steps

Matrix Inverse
Matrix Inverse (Harder)
Inverse (Harder)

Addition (Harder)

Addition (Harder)

Addition (Harder)

Error (Log Scale)

Error (L

Error

Error

- Recurrent
- Programmatic
- Iterative Feedforward
- IREM (Ours)

- Recurrent
- Programmatic
- Iterative Feedforward
- IREM (Ours)

IREM (Ours)

IREM (Ours)

Matrix Completion (Harder)

Matrix Completion (Harder)

Matrix Inverse (Harder)

Matrix Inverse (Harder)

Matrix Inverse (Harder)

Matrix Inverse (Harder)

Matrix Inverse

Matrix Inverse

Inverse (Harder)

Computation Steps

Computation Steps

Computation Steps

Computation Steps

Computation Steps

Computation Steps

Computation Ste

Computation Steps

Error

Error

10^2

10^0

10^1

10^2

10^3

10^{-3}

10^2

10^2

5

0

5

5.0

2.5

0.200

0.200

0.275

0.250

0.225

0.200

0.23

0.22

0.21

0.20

0.23

0.23

0.013

0.013

25

0.200

0

10

20

0

10

0

10

20

0

10

20

0

20

20

Matrix Reasoning

Input A   Input B   →   Prediction   Error Map

Matrix Reasoning

Input A   Input B   →   Prediction   Error Map

Matrix Reasoning

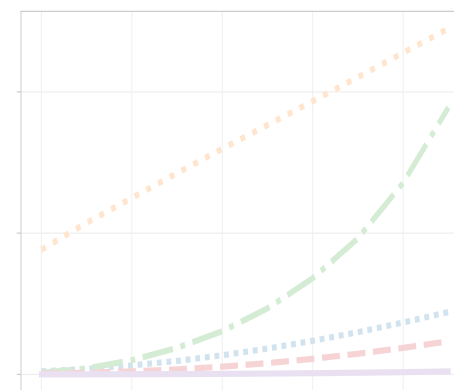Input A   Input B   Prediction   Error Map
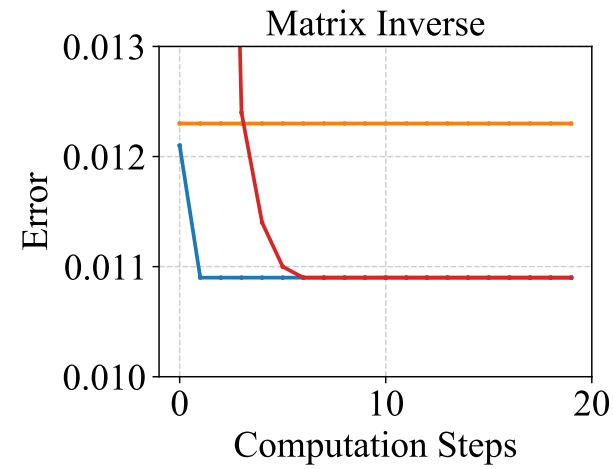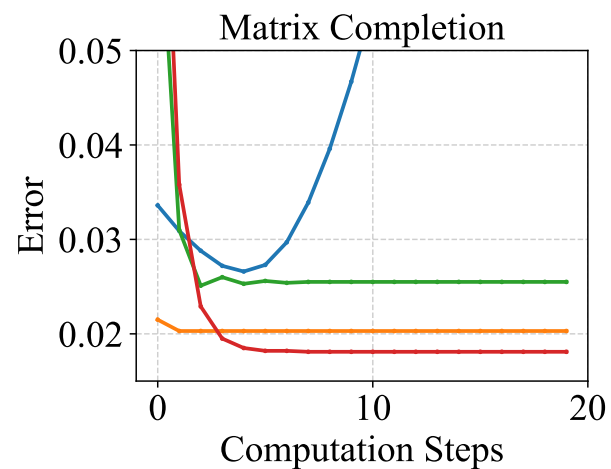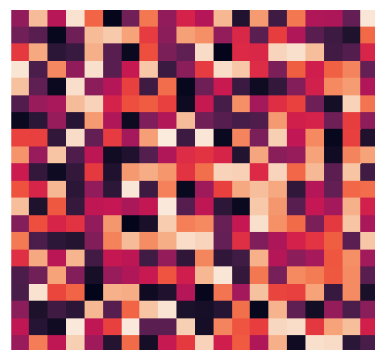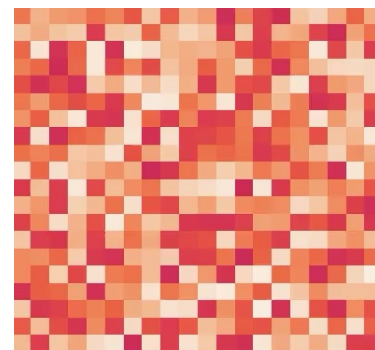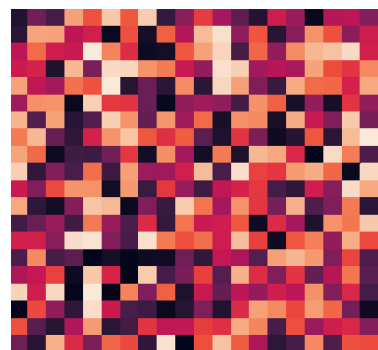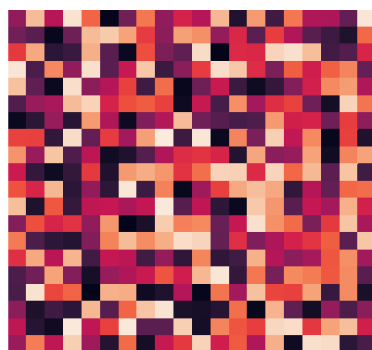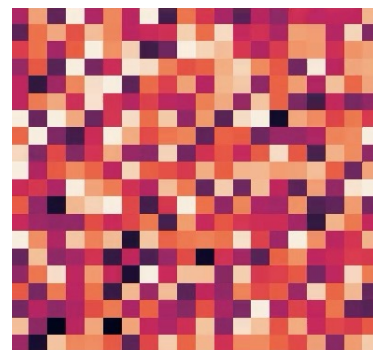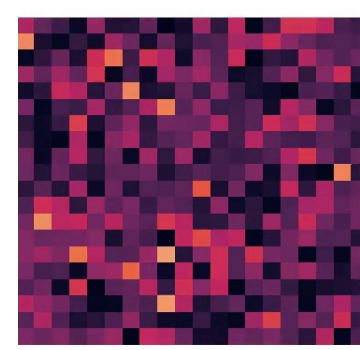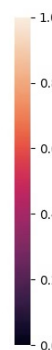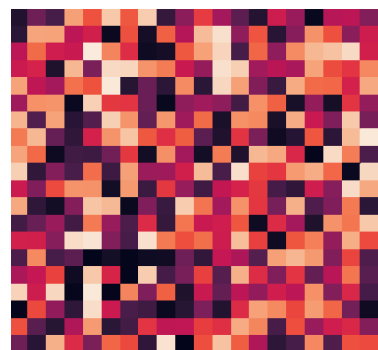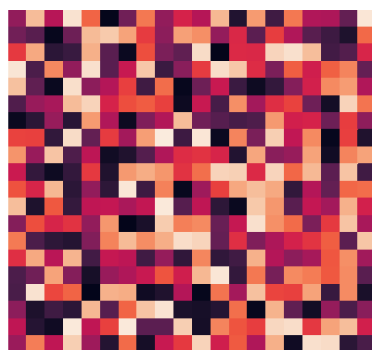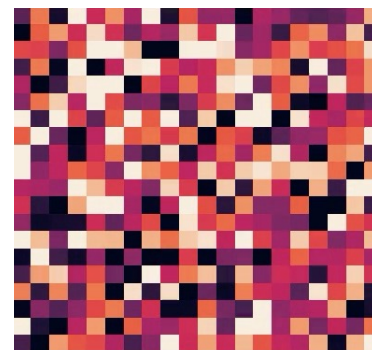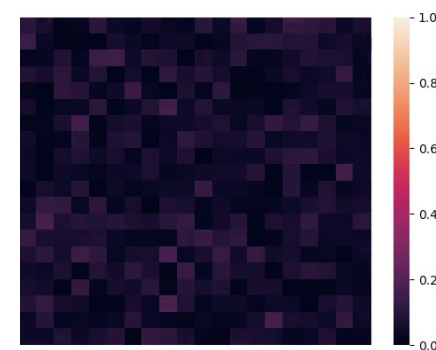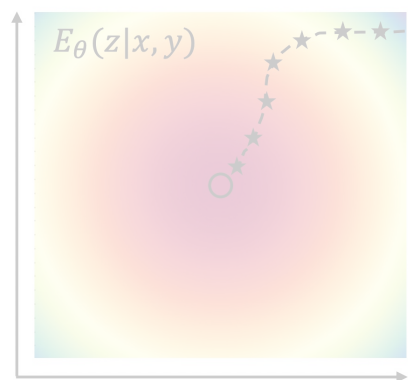
Reasoning as Energy
Minimization

$E_\theta(z|x,y)$

Graph
Reasoning

Matrix
Reasoning

Recursive
Reasoning

Recursively Nesting Applied Factors
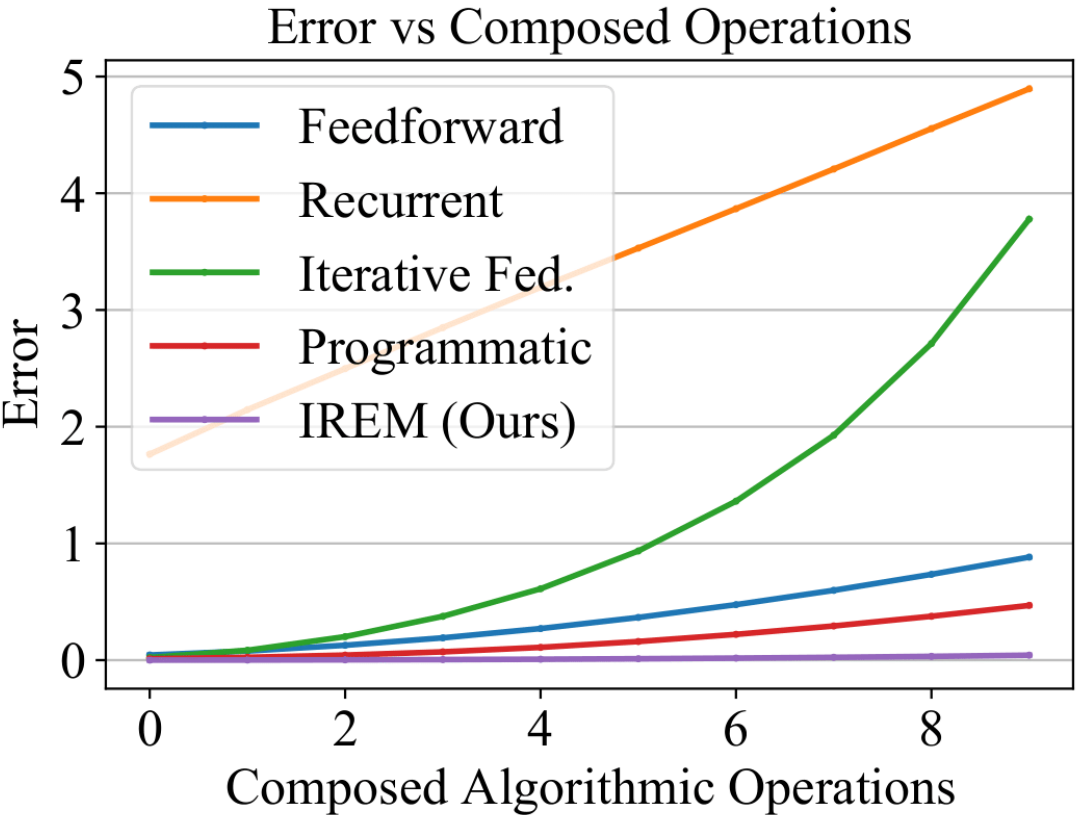
Error vs Composed Operations

Qualitative Visualization

$$\begin{bmatrix} -0.695 & -0.840 & 0.691 \\ 0.488 & -0.961 & -0.048 \\ -0.824 & -0.235 & 0.162 \end{bmatrix} + \begin{bmatrix} -0.590 & 0.972 & -0.544 \\ -0.318 & 0.748 & 0.638 \\ -0.216 & -0.118 & 0.864 \end{bmatrix} +$$

$$\begin{bmatrix} -0.676 & -0.688 & 0.422 \\ 0.075 & 0.172 & -0.963 \\ 0.698 & 0.837 & -0.735 \end{bmatrix} + \begin{bmatrix} 0.945 & -0.432 & -0.940 \\ -0.567 & 0.989 & -0.302 \\ -0.526 & 0.492 & 0.574 \end{bmatrix} = \begin{bmatrix} -1.033 & -1.076 & -0.433 \\ -0.319 & 1.010 & -0.677 \\ -0.867 & 0.984 & 0.843 \end{bmatrix}$$

Qualitative Visualization

$$\begin{bmatrix} -0.695 & -0.840 & 0.691 \\ 0.488 & -0.961 & -0.048 \\ -0.824 & -0.235 & 0.162 \end{bmatrix} + \begin{bmatrix} -0.590 & 0.972 & -0.544 \\ -0.318 & 0.748 & 0.638 \\ -0.216 & -0.118 & 0.864 \end{bmatrix} +$$

$$\begin{bmatrix} -0.676 & -0.688 & 0.422 \\ 0.075 & 0.172 & -0.963 \\ 0.698 & 0.837 & -0.735 \end{bmatrix} + \begin{bmatrix} 0.945 & -0.432 & -0.940 \\ -0.567 & 0.989 & -0.302 \\ -0.526 & 0.492 & 0.574 \end{bmatrix} = \begin{bmatrix} -1.033 & -1.076 & -0.433 \\ -0.319 & 1.010 & -0.677 \\ -0.867 & 0.984 & 0.843 \end{bmatrix}$$

Ground Truth Answer
$$\begin{bmatrix} -1.016 & -0.988 & -0.371 \\ -0.322 & 0.948 & -0.675 \\ -0.868 & 0.976 & 0.865 \end{bmatrix}$$

Qualitative Visualization

$$\begin{bmatrix} -0.695 & -0.840 & 0.691 \\ 0.488 & -0.961 & -0.048 \\ -0.824 & -0.235 & 0.162 \end{bmatrix} + \begin{bmatrix} -0.590 & 0.972 & -0.544 \\ -0.318 & 0.748 & 0.638 \\ -0.216 & -0.118 & 0.864 \end{bmatrix} +$$

$$\begin{bmatrix} -0.676 & -0.688 & 0.422 \\ 0.075 & 0.172 & -0.963 \\ 0.698 & 0.837 & -0.735 \end{bmatrix} + \begin{bmatrix} 0.945 & -0.432 & -0.940 \\ -0.567 & 0.989 & -0.302 \\ -0.526 & 0.492 & 0.574 \end{bmatrix} = \begin{bmatrix} -1.033 & -1.076 & -0.433 \\ -0.319 & 1.010 & -0.677 \\ -0.867 & 0.984 & 0.843 \end{bmatrix}$$
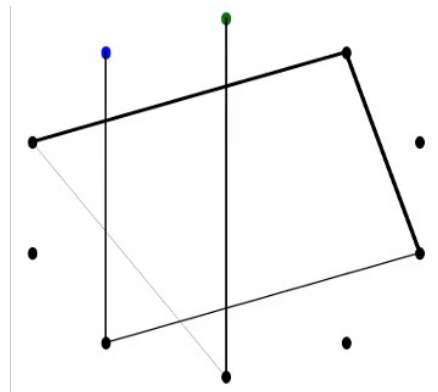
Ground Truth Answer $\begin{bmatrix} -1.016 & -0.988 & -0.371 \\ -0.322 & 0.948 & -0.675 \\ -0.868 & 0.976 & 0.865 \end{bmatrix}$
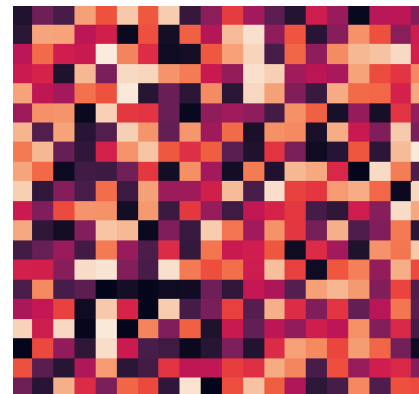
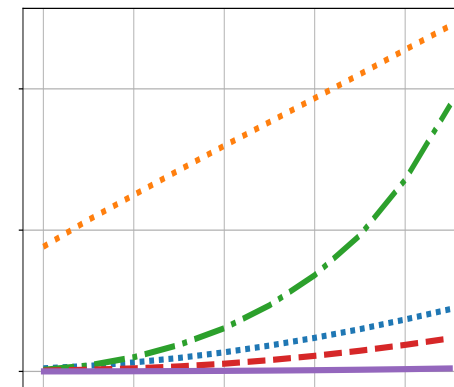First nine entries in 400 matrix
outputs!

Reasoning as Energy Minimization

$E_\theta(z|x,y)$
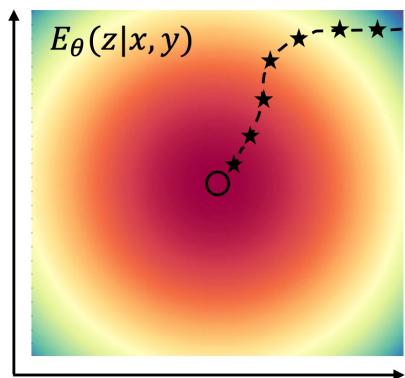
Graph Reasoning

Matrix Reasoning

Recursive Reasoning

Learning Iterative Reasoning through Energy Minimization

Yilun Du[1], Shuang Li[1], Joshua B. Tenenbaum[1,2], Igor Mordatch[3]

[1]MIT CSAIL, [2]MIT BCS, [3]Google Brain