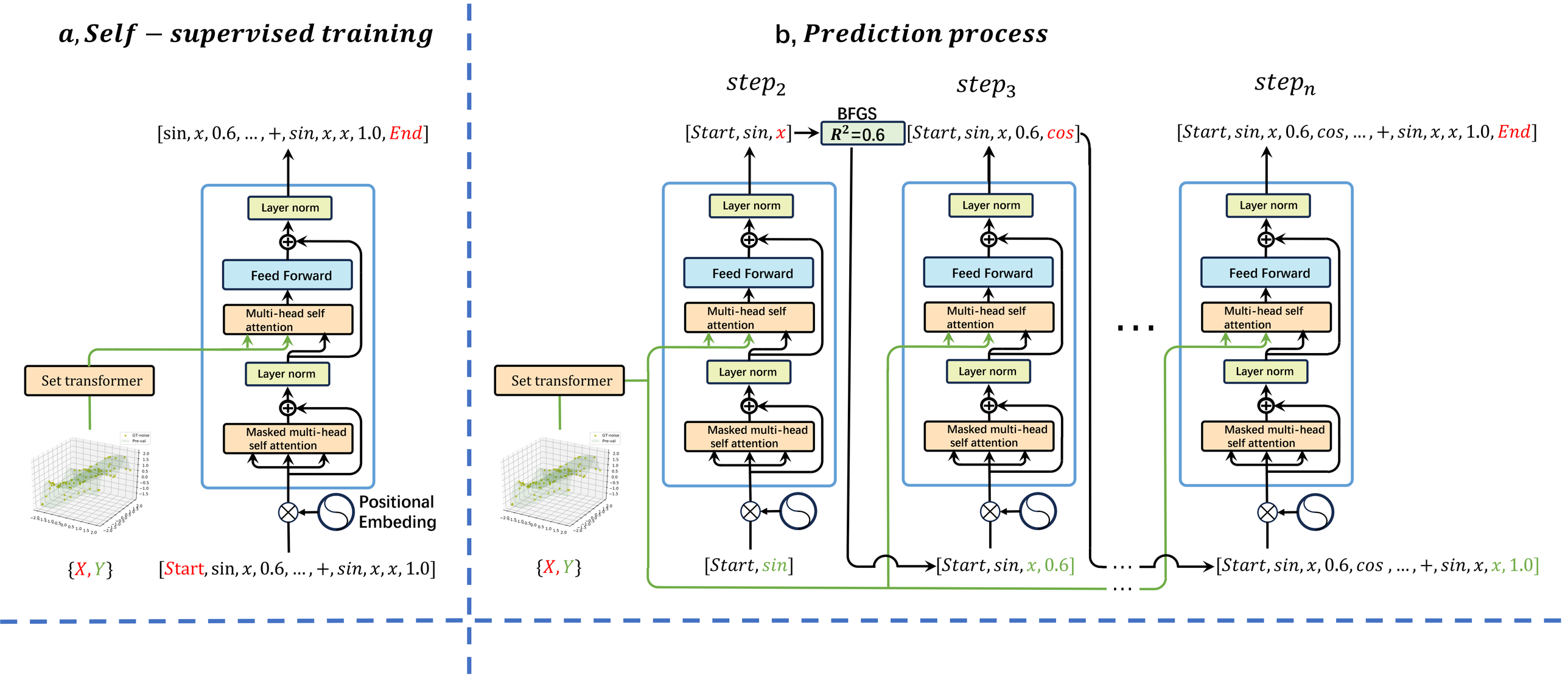
a, Self − supervised training
[sin, x, 0.6, … , +, sin, x, x, 1.0, End]
Layer norm
Feed Forward
Multi-head self attention
Layer norm
Masked multi-head self attention
Set transformer
Positional Embeding
{X, Y}
[Start, sin, x, 0.6, … , +, sin, x, x, 1.0]
b, Prediction process
step2
step3
stepn
BFGS
R²=0.6
[Start, sin, x]
[Start, sin, x, 0.6, cos]
[Start, sin, x, 0.6, cos, … , +, sin, x, x, 1.0, End]
Layer norm
Feed Forward
Multi-head self attention
Layer norm
Masked multi-head self attention
Set transformer
{X, Y}
[Start, sin]
[Start, sin, x, 0.6]
[Start, sin, x, 0.6, cos , … , +, sin, x, x, 1.0]