

# Second-Order Forward mode Optimization of RNNs for neuroscience

(SOFO)



Youjing  
Yu



Rui  
Xia



Qingxi  
Ma



Máté  
Lengyel

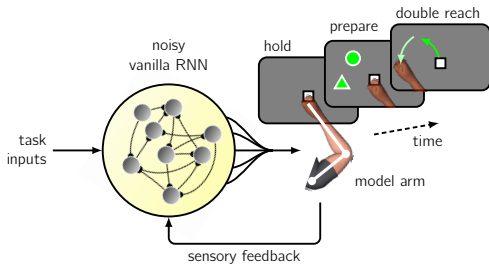


Guillaume  
Hennequin

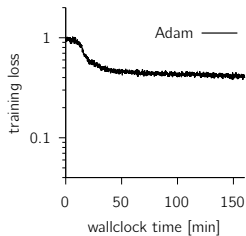
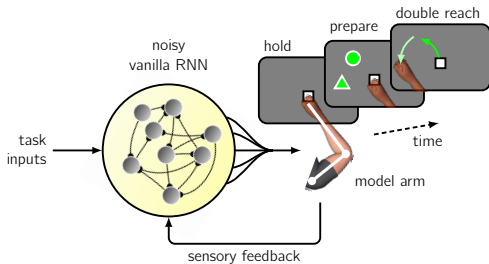


UNIVERSITY OF  
CAMBRIDGE

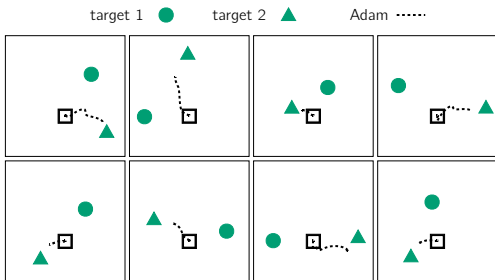
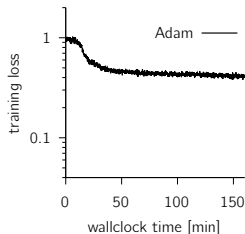
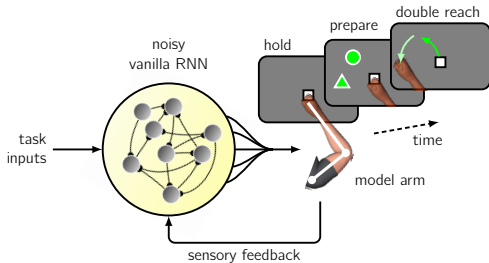
# Training RNNs on neuroscience tasks is difficult



# Training RNNs on neuroscience tasks is difficult

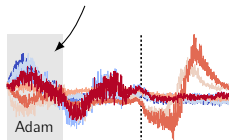


# Training RNNs on neuroscience tasks is difficult



**shoulder torque**  
prepare reach 1 reach 2

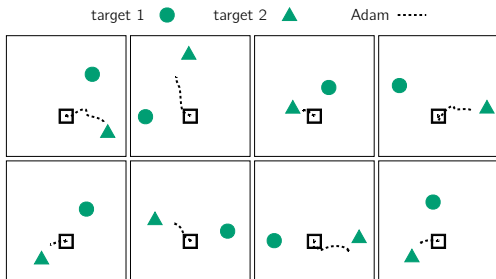
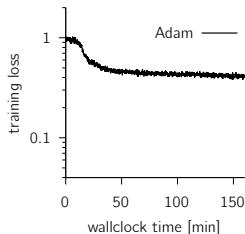
Adam fails to train networks to delay movement as instructed, and RNN fails to double-reach



# Training RNNs on neuroscience tasks is difficult

Here, we develop **SOFO**:

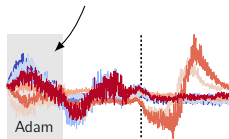
- a 2nd-order optimizer



## shoulder torque

prepare reach 1 reach 2

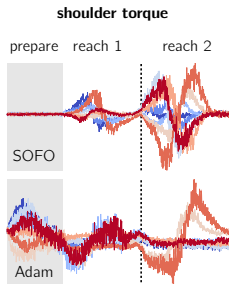
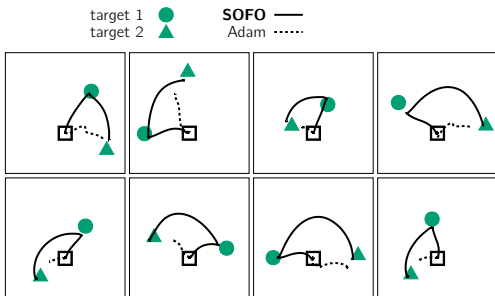
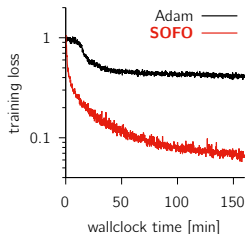
Adam fails to train networks to delay movement as instructed, and RNN fails to double-reach



# Training RNNs on neuroscience tasks is difficult

Here, we develop **SOFO**:

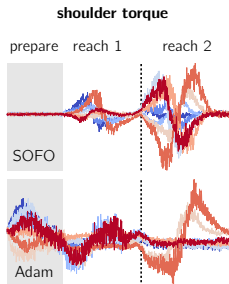
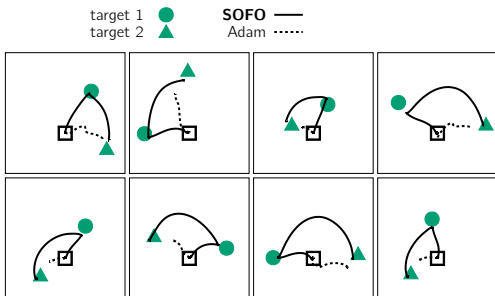
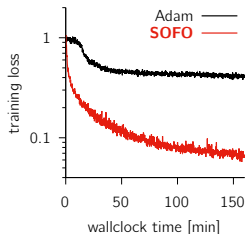
- a **2nd-order optimizer**, which
- greatly **facilitates RNN training**



# Training RNNs on neuroscience tasks is difficult

Here, we develop **SOFO**:

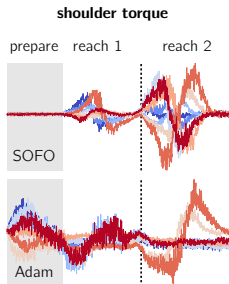
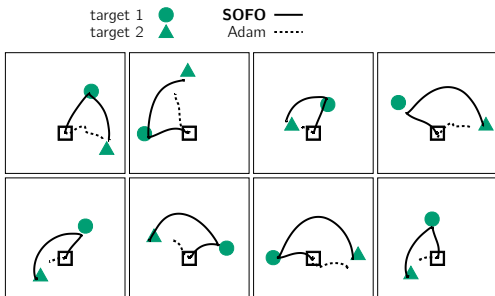
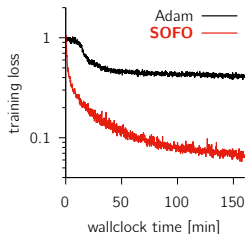
- a **2nd-order optimizer**, which
- greatly **facilitates RNN training**, whilst
- **not** requiring backprop



# Training RNNs on neuroscience tasks is difficult

Here, we develop **SOFO**:

- a **2nd-order optimizer**, which
- greatly **facilitates RNN training**, whilst
- **not** requiring backprop, thus having
- **small memory complexity** independent of time horizon





# SOFO algorithm

$\theta_t$



# SOFO algorithm

$\theta^*$



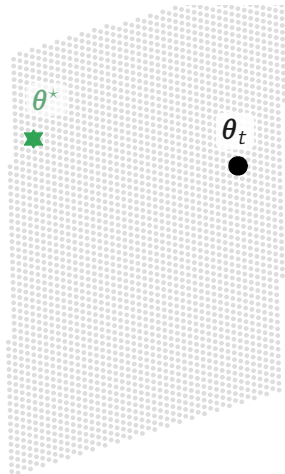
$\theta_t$



# SOFO algorithm

random subspace

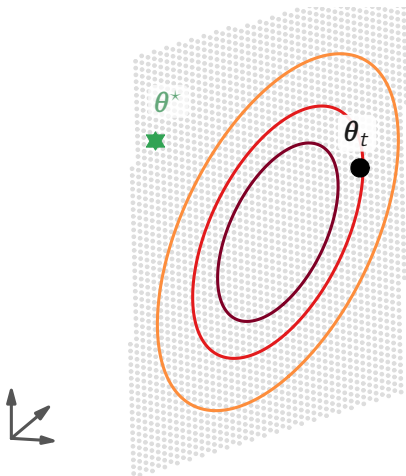
$\Theta_t$



# SOFO algorithm

random subspace

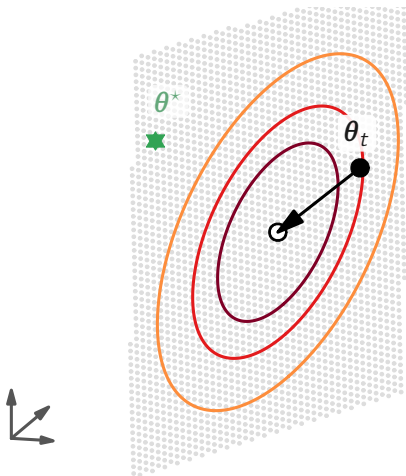
$\Theta_t$



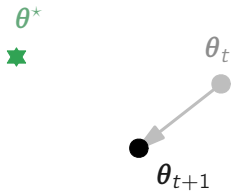
# SOFO algorithm

random subspace

$\Theta_t$

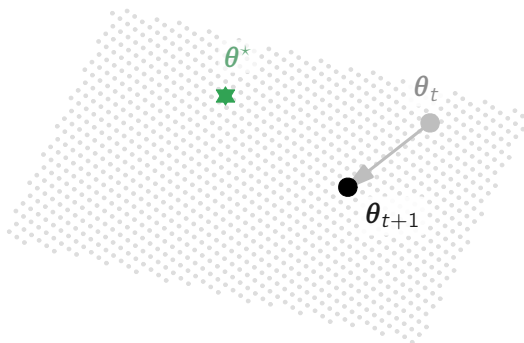


# SOFO algorithm



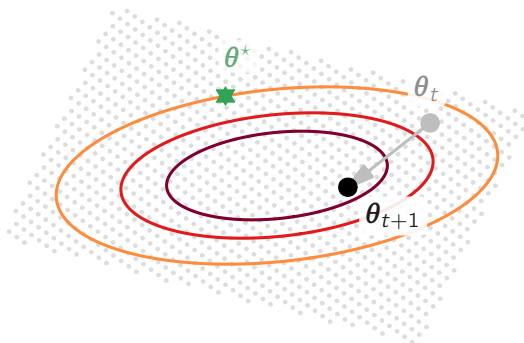
# SOFO algorithm

another random  
subspace  $\Theta_{t+1}$



# SOFO algorithm

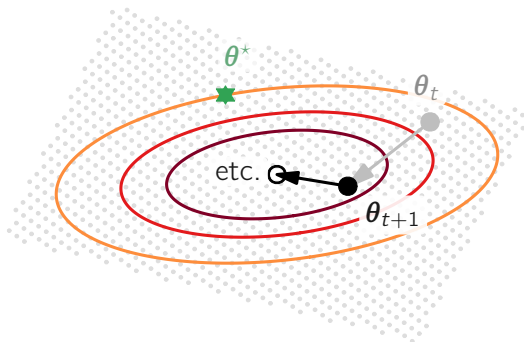
another random  
subspace  $\Theta_{t+1}$





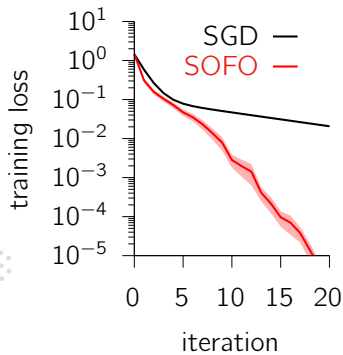
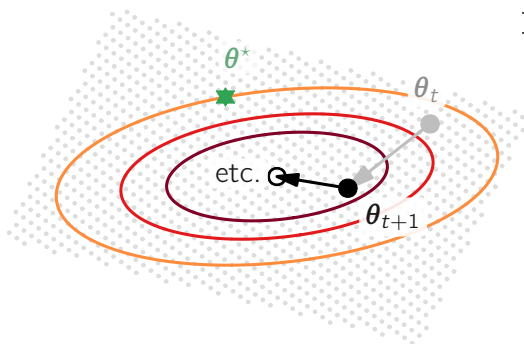
# SOFO algorithm

another random  
subspace  $\Theta_{t+1}$



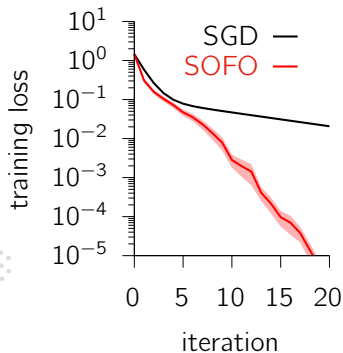
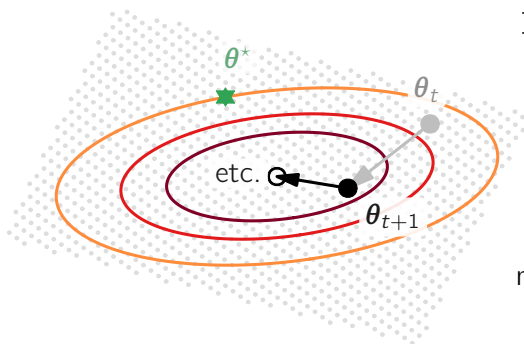
# SOFO algorithm

another random  
subspace  $\Theta_{t+1}$



# SOFO algorithm

another random  
subspace  $\Theta_{t+1}$

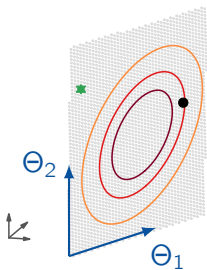


$$\text{need } \tilde{\mathbf{g}} = \Theta^T \mathbf{g}$$

$$\text{and } \tilde{\mathbf{G}} = \Theta^T \mathbf{G} \Theta$$

both can be obtained *without backpropagation*

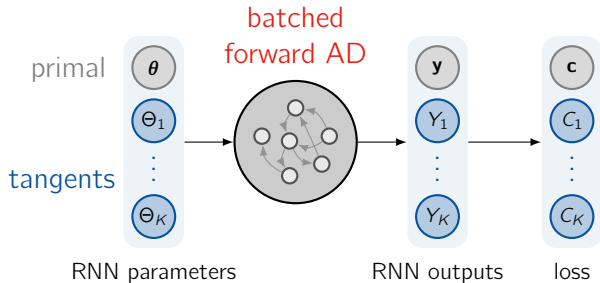
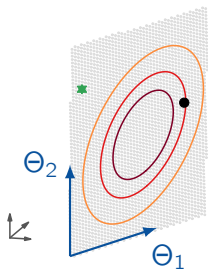
# SOFO algorithm



need  $\tilde{\mathbf{g}} = \Theta^T \mathbf{g}$

and  $\tilde{\mathbf{G}} = \Theta^T \mathbf{G} \Theta$

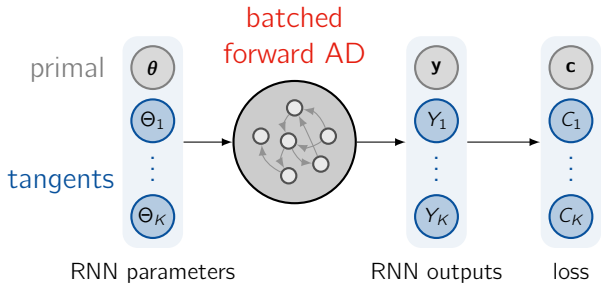
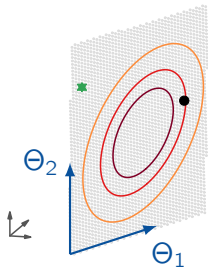
# SOFO algorithm



$$\text{need } \tilde{\mathbf{g}} = \Theta^T \mathbf{g}$$

$$\text{and } \tilde{\mathbf{G}} = \Theta^T \mathbf{G} \Theta$$

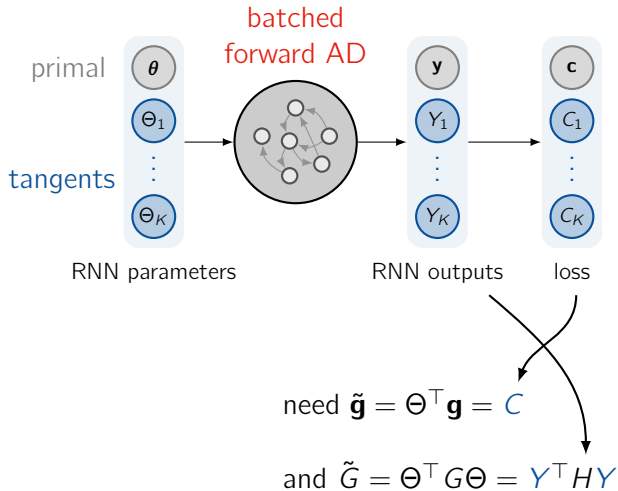
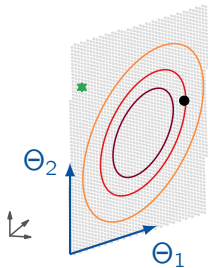
# SOFO algorithm



need  $\tilde{\mathbf{g}} = \Theta^T \mathbf{g} = \mathbf{C}$

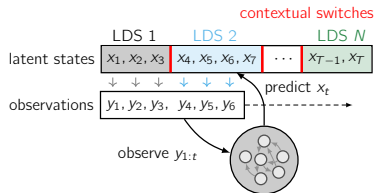
and  $\tilde{\mathbf{G}} = \Theta^T \mathbf{G} \Theta$

# SOFO algorithm



# SOFO is memory and compute efficient

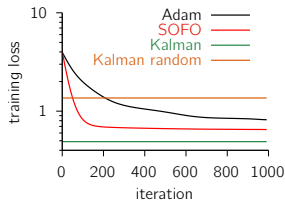
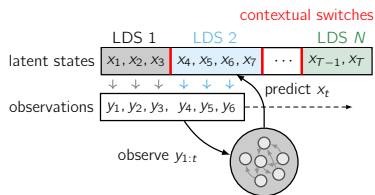
training a vanilla RNN to learn adaptive Kalman filtering





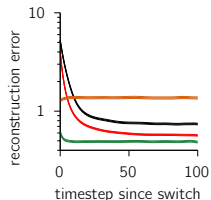
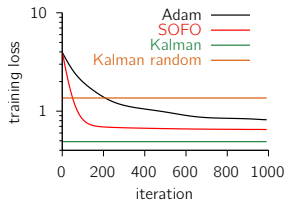
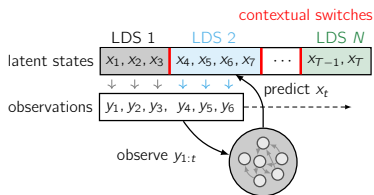
# SOFO is memory and compute efficient

training a vanilla RNN to learn adaptive Kalman filtering



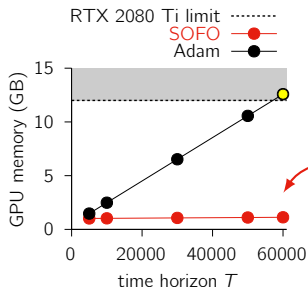
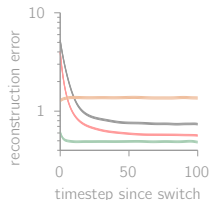
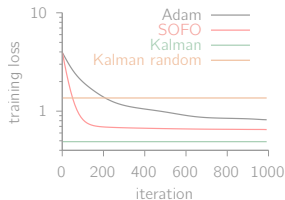
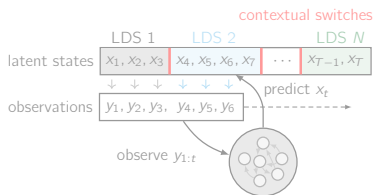
# SOFO is memory and compute efficient

training a vanilla RNN to learn adaptive Kalman filtering



# SOFO is memory and compute efficient

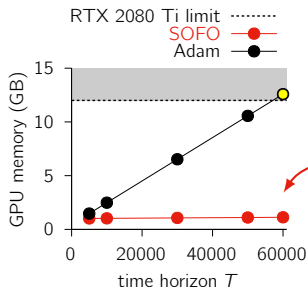
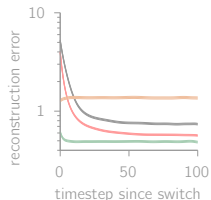
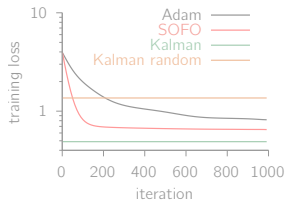
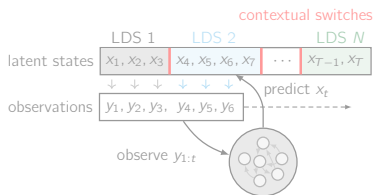
training a vanilla RNN to learn adaptive Kalman filtering



SOFO can handle arbitrarily long time horizons

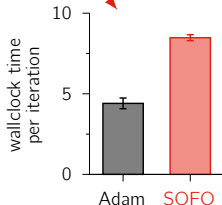
# SOFO is memory and compute efficient

training a vanilla RNN to learn adaptive Kalman filtering



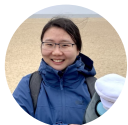
SOFO exploits GPU parallelism otherwise underused by sequential models

SOFO can handle arbitrarily long time horizons



# Second-Order Forward mode Optimization of RNNs for neuroscience

(SOFO)



Youjing  
Yu



Rui  
Xia



Qingxi  
Ma



Máté  
Lengyel



Guillaume  
Hennequin



UNIVERSITY OF  
CAMBRIDGE