A Hybrid Gradient Algorithm for Linear Regression with Hybrid Signals

A. Saoud[1], M. Maghenem[2], and R. G. Sanfelice[3]

[1]Department of Electrical and Computer Engineering, University of California, Berkeley

[2]CNRS, Gipsa lab, Grenoble, France

[3]Department of Electrical & Computer Engineering, University of California Santa Cruz

Email: asaoud@berkeley.edu

Review meeting

AFOSR Center of Excellence on Assured Autonomy

Gradient descent

Definition

A linear regression model is of the form

$$y(t) = \theta^{*\top}\phi(t) \qquad t \in \mathbb{R}_{\geq 0} \quad \text{or} \quad t \in \mathbb{N}_{\geq 0}$$

- $y : \mathbb{R}_{\geq 0} \to \mathbb{R}$: the known output
- $\theta^* \in \mathbb{R}^n$: the unknown constant parameter to identify
- $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$: the known regressor

1. Consider the estimator of the form $\hat{y}(t) = \theta^\top(t)\phi(t)$
   - $\hat{y}$ the estimated output.
   - θ the estimate of the unknown parameter.
2. Consider the estimation error

$$e := \hat{y} - y = \tilde{\theta}^\top\phi \qquad \tilde{\theta} := \theta - \theta^*.$$

3. Consider the cost function $J(e) := \frac{1}{2}e^2$.
4. Use the gradient-descent algorithm

Continuous-Time Gradient-Descent Algorithms

We continuously update θ as follows

$$\dot{\theta}(t) = \dot{\tilde{\theta}}(t) = -\nabla_\theta J(e(t)) = -\phi(t)\phi^\top(t)\tilde{\theta}(t).$$

Definition

$\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is continuous-time persistently exciting (PE) if $\exists\, T, \mu > 0$ s.t $\forall\, t_0 \geq 0$, we have

$$\int_{t_0}^{t_0+T} \phi(s)\phi^\top(t)dt \geq \mu I_n.$$

If ϕ is PE and bounded, then the origin for

$$\dot{\tilde{\theta}}(t) = -\phi(t)\phi^\top(t)\tilde{\theta}(t)$$

is uniformly exponentially stable [K. S. Narendra and A. M. Annaswamy. IJC, 87].

Discrete-Time Gradient-Descent Algorithms

We discretely update θ as follows

$$\theta(t+1) = \theta(t) - \nabla_\theta J(e) = \theta(t) - \frac{\phi(t)\phi^\top(t)}{1+|\phi(t)|^2}\tilde{\theta}(t).$$

Definition

$\phi : \mathbb{N}_{\geq 0} \to \mathbb{R}^n$ is Discrete-time PE if $\exists\ J, \mu > 0$ s.t $\forall\ j_0 \geq 0$, we have

$$\sum_{t=j_o}^{j_o+J} \phi(t)\phi(t)^\top \geq \mu I_n.$$

If ϕ is PE and bounded, then the origin for

$$\tilde{\theta}(t+1) = \left[I_n - \frac{\phi(t)\phi^\top(t)}{1+|\phi(t)|^2}\right]\tilde{\theta}(t)$$

is uniformly exponentially stable [G. Tao. John Wiley Sons, 03].

Hybrid Gradient Descent

What if ϕ has both continuous and discrete evolutions?

(e.g., ϕ collects data from a physical system with impacts, a cyber-physical system, or network of systems with discrete and continuous interactions)

Formally, what if ϕ is a hybrid arc?

A hybrid arc ϕ is a function parameterized by hybrid time (t, j):

- Flows parameterized by $t \in \mathbb{R}_{\geq 0} := [0, +\infty)$
- Jumps parameterized by $j \in \mathbb{N}_{\geq 0} := \{0, 1, 2, \ldots\}$

Then, ϕ is defined on a hybrid time-domain of the form
$([0, t_1] \times \{0\}) \cup ([t_1, t_2] \times \{1\}) \cup \ldots ([t_j, t_{j+1}] \times \{j\}) \cup \ldots$

- A hybrid linear regression model is of the form

$$y(t, j) = \theta^{*\top} \phi(t, j) \qquad (t, j) \in \operatorname{dom} \phi = \operatorname{dom} y.$$

Approach: During the continuous evolution of ϕ, we update θ using the continuous gradient descent. During the discrete evolution of ϕ, we update θ using the discrete gradient descent.

Hybrid gradient descent

The hybrid gradient algorithm is defined by the hybrid system

$$\mathcal{H}_g : \begin{cases} \begin{bmatrix} \dot{\tilde{\theta}} \\ \dot{t} \\ \dot{j} \end{bmatrix} = \begin{bmatrix} -\phi(t,j)\phi(t,j)^\top \tilde{\theta} \\ 1 \\ 0 \end{bmatrix} & (\tilde{\theta}, t, j) \in C_g \\ \begin{bmatrix} \tilde{\theta}^+ \\ t^+ \\ j^+ \end{bmatrix} = \begin{bmatrix} \tilde{\theta} - \frac{\phi(t,j)\phi(t,j)^\top}{1+|\phi(t,j)|^2} \tilde{\theta} \\ t \\ j+1 \end{bmatrix} & (\tilde{\theta}, t, j) \in D_g \end{cases}$$

$$D_g := \{(\tilde{\theta}, t, j) \in \mathbb{R}^n \times \operatorname{dom}\phi : (t, j+1) \in \operatorname{dom}\phi\}$$

$$C_g := (\mathbb{R}^n \times \operatorname{dom}\phi) \backslash D_g.$$

Under which conditions we can guarantee uniform exponential stability of the set $\mathcal{A}$?

$$\mathcal{A} := \{(\tilde{\theta}, t, j) \in \mathbb{R}^n \times \operatorname{dom}\phi : \tilde{\theta} = 0\}.$$

Hybrid Persistence of Excitation (hybrid PE)

Definition

A hybrid arc ϕ is hybrid PE if there exist $k > 0$ and $\mu > 0$ such that, for each $(t_o, j_o) \in \operatorname{dom}\phi$ and for each hybrid time domain

$$E := \cup_{j=j_o}^{J} ([t_j, t_{j+1}] \times \{j\}) \subset \operatorname{dom}\phi$$

with $t_{j_o} := t_o$ and $(t_{J+1} - t_o) + (J - j_o) \geq k$, the following holds:

$$\sum_{j=j_o}^{J} \int_{t_j}^{t_{j+1}} \phi(u,j)\phi(u,j)^\top du + \frac{1}{2} \sum_{j=j_o}^{J} \frac{\phi(t_{j+1},j)\phi(t_{j+1},j)^\top}{1 + |\phi(t_{j+1},j)|^2} \geq \mu I_n.$$

- When ϕ is eventually continuous, hybrid PE reduces to continuous PE.
- When ϕ is eventually discrete or Zeno, hybrid PE reduces to discrete PE.

Convergence Rate of the Hybrid Gradient Descent

Theorem: Let ϕ be bounded (by $\bar{\phi}$) and hybrid PE, i.e

$$\sum_{j=j_o}^{J} \int_{t_j}^{t_{j+1}} \phi(u,j)\phi(u,j)^\top du + \frac{1}{2} \sum_{j=j_o}^{J} \frac{\phi(t_{j+1},j)\phi(t_{j+1},j)^\top}{1+|\phi(t_{j+1},j)|^2} \geq \mu I_n.$$

Then, for the hybrid system $\mathcal{H}_g$, representing the hybrid gradient descent algorithm, the set $\mathcal{A} := \{(\tilde{\theta},t,j) \in \mathbb{R}^n \times \operatorname{dom}\phi : \tilde{\theta} = 0\}$ is uniformly exponentially stable, i.e,

$$|\xi(t,j)|_{\mathcal{A}} \leq \kappa e^{-\lambda(t+j)} |\xi(0,0)|_{\mathcal{A}}$$

for any solution ξ to $\mathcal{H}_g$ and for all $(t,j) \in \operatorname{dom}\xi = \operatorname{dom}\phi$, where the convergence rate λ satisfies

$$\lambda := -\frac{\log(1-\alpha)}{k},\ \alpha := \frac{2\mu}{\left(1+(k+2)\sqrt{(\bar{\phi}+2)(1/2+\bar{\phi}(k+1)^2)}\right)^2}.$$

Convergence Rate of the Hybrid Gradient Descent

Sketch of proof

The hybrid gradient algorithm

$$\mathcal{H}_g : \begin{cases} \begin{bmatrix} \dot{\tilde{\theta}} \\ \dot{t} \\ \dot{j} \end{bmatrix} = \begin{bmatrix} -\phi(t,j)\phi(t,j)^\top \tilde{\theta} \\ 1 \\ 0 \end{bmatrix} & (\tilde{\theta}, t, j) \in C_g \\ \begin{bmatrix} \tilde{\theta}^+ \\ t^+ \\ j^+ \end{bmatrix} = \begin{bmatrix} \tilde{\theta} - \frac{\phi(t,j)\phi(t,j)^\top}{1+|\phi(t,j)|^2} \tilde{\theta} \\ t \\ j+1 \end{bmatrix} & (\tilde{\theta}, t, j) \in D_g \end{cases}$$

Convergence Rate of the Hybrid Gradient Descent

Sketch of proof

The hybrid gradient algorithm is defined by the hybrid system

$$\mathcal{H}_g : \begin{cases} \begin{bmatrix} \dot{\tilde{\theta}} \\ \dot{t} \\ \dot{j} \end{bmatrix} = \begin{bmatrix} -A(t,j)\tilde{\theta} \\ 1 \\ 0 \end{bmatrix} & (\tilde{\theta}, t, j) \in C_g \\ \begin{bmatrix} \tilde{\theta}^+ \\ t^+ \\ j^+ \end{bmatrix} = \begin{bmatrix} \tilde{\theta} - B(t,j)\tilde{\theta} \\ t \\ j+1 \end{bmatrix} & (\tilde{\theta}, t, j) \in D_g \end{cases}$$

Structural properties of the matrices A and B:

1. For each $(t,j) \in \operatorname{dom} A = \operatorname{dom} B = \mathcal{S}$,
   $A(t,j) = A(t,j)^\top \geq 0 \qquad B(t,j) = B(t,j)^\top \geq 0$
2. For each $(t,j) \in \mathcal{S}$, $|B(t,j)| \leq 1$
3. There exists $\bar{A} > 0$ such that
   $\operatorname{ess\,sup}\{|A(t,j)| : (t,j) \in \mathcal{S}\} \leq \bar{A}$.

Convergence Rate of the Hybrid Gradient Descent

Sketch of proof: uniform stability

We consider the Lyapunov function candidate

$$V(x) := \frac{1}{2}\tilde{\theta}^\top \tilde{\theta} = \frac{1}{2}|x|_{\mathcal{A}}. \qquad (1)$$

Now for each $x \in C$, we have

$$\begin{aligned} \langle \nabla V(x), F(x) \rangle &= \dot{x}(t,j)^\top x(t,j) \\ &= -x(t,j)^\top A(t,j) x(t,j) \leq 0. \end{aligned}$$

For each $(t,j) \in \mathcal{S}$ such that $(t,j+1) \in \mathcal{S}$,

$$\begin{aligned} V(x(t,j+1)) - V(x(t,j)) &= V(G(x(t,j)) - V(x(t,j)) \\ &= -\frac{1}{2} x(t,j)^\top B(t,j) x(t,j) \leq 0. \end{aligned}$$

Hence, for each maximal solution x to $\mathcal{H}$, we conclude that

$$|x(t,j)|_{\mathcal{A}} \leq |x(0,0)|_{\mathcal{A}} \qquad \forall (t,j) \in \operatorname{dom} x,$$

which concludes uniform stability of the set $\mathcal{A}$.

Convergence Rate of the Hybrid Gradient Descent

Sketch of proof: exponential attractivity (1)

To show exponential stability of the closed set $\mathcal{A}$, we show that, for each $(t_o, j_o) \in \mathcal{S}$ and for each hybrid domain

$$E := \cup_{j=j_o}^{J} ([t_j, t_{j+1}] \times \{j\}) \subset \mathcal{S}$$

with $(t_{J+1} - t_{j_o}) + (J - j_o) \geq k$, the following holds

$$V(x(t_{J+1}, J)) - V(x(t_o, j_o)) \leq (1 - \alpha)V(x(t_o, j_o)), \qquad (2)$$

To prove (2), we note that

$$\tilde{V} := V(x(t_{J+1}, J)) - V(x(t_o, j_o)) = \sum_{j=j_o}^{J} [V_F(t_j, t_{j+1}, j) + V_G(t_{j+1}, j, j +$$

where

$$V_F(t_j, t_{j+1}, j) := V(x(t_{j+1}, j)) - V(x(t_j, j))$$

$$V_G(t_{j+1}, j, j+1) := V(x(t_{j+1}, j+1)) - V(x(t_{j+1}, j)).$$

Convergence Rate of the Hybrid Gradient Descent

Sketch of proof: exponential attractivity (2)

$$\tilde{V} := V(x(t_{J+1}, J)) - V(x(t_o, j_o)) = \sum_{j=j_o}^{J} [V_F(t_j, t_{j+1}, j) + V_G(t_{j+1}, j, j +$$

Lemma (Variation during flow intervals)

For each $\rho > 0$, the function V_F satisfies

$$V_F(t_j, t_{j+1}, j) \leq -\rho \bar{A}(\bar{A} + 2)(2(j - j_o) + 1)(t_{j+1} - t_j)^2 \tilde{V} - \frac{\rho}{1+\rho} \int_{t_j}^{t_{j+1}} \left| A(s, j)^{\frac{1}{2}} x(t_o, j_o) \right|^2 ds,$$

Convergence Rate of the Hybrid Gradient Descent

Sketch of proof: exponential attractivity (2)

$$\tilde{V} := V(x(t_{J+1}, J)) - V(x(t_o, j_o)) = \sum_{j=j_o}^{J} [V_F(t_j, t_{j+1}, j) + V_G(t_{j+1}, j, j +$$

Lemma (Variation at jumps)

For each $\rho > 0$, the function V_G satisfies

$$\begin{aligned} V_G(t_{j+1}, j, j+1) \leq & \frac{1}{2}\rho(2(j - j_o) + 1)(\bar{A} + 2)\tilde{V} \\ & - \frac{1}{2}\frac{\rho}{1+\rho}\left|B(t_{j+1}, j)^{\frac{1}{2}} x(t_o, j_o)\right|^2, \end{aligned}$$

The result holds by combining the two Lemmas

Numerical Example

Consider the hybrid linear regression model with

$$\phi(t,j) := \begin{cases} [\sin(t) \quad 0]^\top & \text{if} \quad t \in (2j\pi, 2(j+1)\pi), \quad j \in \mathbb{N} \\ [0.5 \quad 1]^\top & \text{otherwise.} \end{cases}$$

- The continuous-time evolution of ϕ not continuous-time PE since, for each $t_o > 0$ and $T > 0$, we have

$$\int_{t_o}^{t_o+T} \phi(s,j)\phi(s,j)^\top ds = \begin{bmatrix} \int_{t_o}^{t_o+T} \sin(s)^2 ds & 0 \\ 0 & 0 \end{bmatrix}.$$

- The discrete-time evolution of ϕ is not discrete-time PE since the matrix

$$\phi(2j\pi,j)\phi(2j\pi,j)^\top := \begin{bmatrix} 0.25 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad j \in \mathbb{N},$$

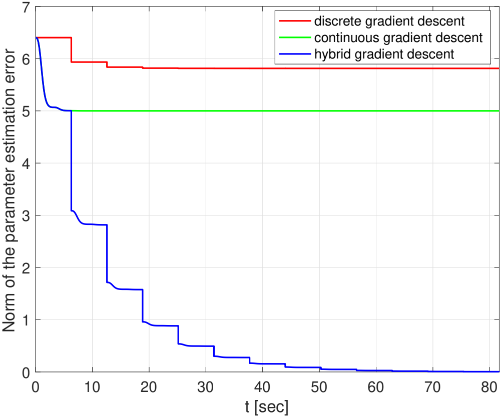
is constant and not full rank.

- However, ϕ is hybrid PE with $k = 2\pi + 1$ and $\mu = 0.21$.

Numerical example

Consider the hybrid linear regression model with

$$\begin{cases} \phi(t,j) = [\sin(t) \quad 0]^\top & \text{if} \quad t \in (2j\pi, 2(j+1)\pi) \\ \phi(t,j) = [-0.5 \quad 1]^\top & \text{if} \quad t = 2j\pi \end{cases}$$



Conclusion and future work

- A hybrid gradient algorithm for linear regression model is proposed;
- Uniform exponential stability is proved under a hybrid PE condition;
- Exploring new approaches to check the hybrid PE;
- A first step towards a general hybrid adaptive-control framework, which is more suitable for cyber-physical systems, exhibiting both continuous and discrete evolutions.

This research has been partially supported by NSF Grants no. ECS-1710621, CNS-1544396, and CNS-2039054, by AFOSR Grants no. FA9550-19-1-0053, FA9550-19-1-0169, and FA9550-20-1-0238, and by CITRIS and the Banatao Institute at the University of California.