Fast Gradient-Descent Methods for Temporal-Difference Learning with Linear Function Approximation

Adarsh Shah
Apoorv Pandey
Bhartendu Kumar

Indian Institute Of Science

19473, adarshshah@iisc.ac.in
19598, apoorvpandey@iisc.ac.in
19649, bhartendukumar@iisc.ac.in

April 15, 2022

TD Learning

- Not True Gradient-Descent method :
  More conditions / restrictions
  Less Robust
- Off- Policy might not converge at all :
  Intuitively: Estimating value of different policy than the one governing the MDP might be unstable
- Why ONLY Gradient-Descent methods?
  Complexity: Second Order Methods are generally $\mathcal{O}(n^2)$
  While gradient based ones are $\mathcal{O}(n)$
  LSTD by Bradtke & Barto and extended by Boyan approximate the fixed point of TD with $\mathcal{O}(n^2)$ per iteration as that has to update Matrix whereas TD(λ) updates the feature vector at each time with a complexity of $\mathcal{O}(n)$

Objective Functions

Mean Square Bellman Error

- No need to approximate V^* in gadient
- Can Alternatively look as gradient descent solving for fixed point of operator T
- Averaged over state space proportional to the time Markov Chain is in that state
- $\text{MSBE}(\theta) = ||V_\theta - TV_\theta||_D^2$
- TD,LSTD,GTD and many of the temporal difference algorithms do not converge to the minima of this objective function

Issue: In Linear Settings, V_θ will always lie in the column space of Φ but TV_θ might potentially lie out of that subspace

Objective Functions

Norm of the Expected TD update

- $NEU(\theta) = \mathbb{E}[\delta\phi]^T \, \mathbb{E}[\delta\phi]$
- $\mathbb{E}[\delta\phi]$ can be viewed as an error in the current solution θ
- $NEU(\theta)$ is a measure of how far we are away from the TD solution

This is the objective function for the GTD algorithm.

Mean Square Projected Bellman Error

- Π : Projection operator that projects any vector into column space of Φ
- $\Pi = \Phi(\Phi^T D \Phi)^{-1} \Phi^T D$
- Both the vectors V_θ and $\Pi T V_\theta$ are in same subspace
- $\text{MSPBE}(\theta) = ||V_\theta - \Pi T V_\theta||_D^2$

Choice: This is the objective function that will be minimized by updating θ iteratively

Objective Functions: MSBE vs MSPBE

- Under (Linear) Function Approximation MSBE is just guided by TD error minimization
- But MSPBE sees that the value it has approximated will be again projected into the same linear space.

MSPBE as Expectation Formulation

$$
\begin{aligned}
\Pi^T D \Pi &= (\Phi(\Phi^T D\Phi)^{-1}\Phi^T D)^T D(\Phi(\Phi^T D\Phi)^{-1}\Phi^T D) \\
&= D^T\Phi(\Phi^T D\Phi)^{-1}\Phi^T D(\Phi(\Phi^T D\Phi)^{-1}\Phi^T D) \\
&= D^T\Phi(\Phi^T D\Phi)^{-1}\Phi^T D
\end{aligned}
$$

$$\mathbb{E}[\delta\phi] = \Phi^T D(TV_\theta - V_\theta)$$

$$\mathbb{E}[\phi\phi^T] = \Phi^T D\Phi$$

$$
\begin{aligned}
\mathrm{MSPBE}(\theta) &= ||V_\theta - \Pi TV_\theta||_D^2 \\
&= ||\Pi(V_\theta - TV_\theta)||_D^2 \\
&= (\Pi(V_\theta - TV_\theta))^T D\Pi(V_\theta - TV_\theta) \\
&= (V_\theta - TV_\theta)^T \Pi^T D\Pi(V_\theta - TV_\theta) \\
&= (V_\theta - TV_\theta)^T D^T\Phi(\Phi^T D\Phi)^{-1}\Phi^T D(V_\theta - TV_\theta) \\
&= \mathbb{E}[\delta\phi]^T \; \mathbb{E}[\phi\phi^T]^{-1} \; \mathbb{E}[\delta\phi]
\end{aligned}
$$

MSPBE as an improved Objective Function (over NEU)

- As $\mathbb{E}[\phi\phi^T] = \Phi^T D \Phi$, thus if Φ has full column rank and the State Markov Chain of the behaviour policy has unique stationary distribution. Then $\mathbb{E}[\phi\phi^T]^{-1}$ is Positive Definite.
- Thus here also like NEU (GTD), the TD solution is that value of θ where $\mathbb{E}[\delta\phi]$ is 0.
- Hence we are tracking the same minima but we can see the objective as norm square of $\mathbb{E}[\delta\phi]$ with respect to Matrix norm of $\mathbb{E}[\phi\phi^T]^{-1}$.
- Gradient Descent on MSPBE is like Scaled Gradient Descent of NEU.
$\nabla_\theta MSPBE(\theta) = -2\,\mathbb{E}[(\phi - \gamma\phi')\phi^T]\ \mathbb{E}[\phi\phi^T]^{-1}\ \mathbb{E}[\delta\phi]$
$\nabla_\theta NEU(\theta) = -2\,\mathbb{E}[(\phi - \gamma\phi')\phi^T]\ \mathbb{E}[\delta\phi]$
- Based on the empirical results GTD2 i.e. the SGD w.r.t. MSPBE is faster than GTD i.e. the SGD w.r.t. NEU. Thus this Gradient Scaling leads to faster convergence.

Second Parameter

- Gradient of MSPBE
  $\nabla_\theta MSPBE(\theta) = -2\,\mathbb{E}[(\phi - \gamma\phi')\phi^T]\ \mathbb{E}[\phi\phi^T]^{-1}\ \mathbb{E}[\delta\phi]$

Using Second Parameter w

- $w \simeq \mathbb{E}[\phi\phi^T]^{-1}\ \mathbb{E}[\delta\phi]$

Another parameter w is introduced to solve the following problems:

- As 2 independent Expectation terms of gradient $\nabla_\theta MSPBE(\theta)$ is approximated by w which is updated in each iteration, we just need one sample to observe to be used in the first expectation in Gradient.
- Here $V_\theta = \Phi\theta$,thus w is a justifiable approximation, i.e. near to actual value.

- Gradient of MSPBE
  $\nabla_\theta MSPBE(\theta) \simeq -2\,\mathbb{E}[(\phi - \gamma\phi')\phi^T]\ w$

Deriving GTD2

- Gradient of MSPBE
  $\nabla_\theta MSPBE(\theta) \simeq -2\,\mathbb{E}[(\phi - \gamma\phi')\phi^T]\ w$
- Gradient descent Step of MSPBE(θ) :
  $\theta_{k+1} = \theta_k - \alpha_k(-2\,\mathbb{E}[(\phi_k - \gamma\phi'_k)\phi_k^T]\ w_K)$

  $\mathbb{E}[(\phi_k - \gamma\phi'_k)(\phi_k^T)]$ can be substituted with $(\phi_k - \gamma\phi'_k)(\phi_k^T)$ to get the SGD updates.
- Stochastic Gradient descent Step of MSPBE(θ) :

  $\theta_{k+1} = \theta_k + \alpha_k(\phi_k - \gamma\phi'_k)(\phi_k^T w_k)$
  w_k update:
  $w_{k+1} = w_k + \beta_k(\phi_k - \gamma\phi'_k)(\phi_k^T w_k)$
- Per iteration Complexity : $\mathcal{O}(n)$
  n: dimension of linear approximating vector θ

Deriving TDC

- Gradient of MSPBE

$$
\begin{aligned}
\frac{-1}{2}\nabla_\theta MSPBE(\theta) &= \mathbb{E}[(\phi - \gamma\phi')\phi^T]\ \mathbb{E}[\phi\phi^T]^{-1}\ \mathbb{E}[\delta\phi] \\
&= (\mathbb{E}[(\phi\phi^T] - \gamma\,\mathbb{E}[\phi'\phi^T])\ \mathbb{E}[\phi\phi^T]^{-1}\ \mathbb{E}[\delta\phi] \\
&= \mathbb{E}[\delta\phi] - \gamma\,\mathbb{E}[\phi'\phi^T]\ \mathbb{E}[\phi\phi^T]^{-1}\ \mathbb{E}[\delta\phi] \\
&\simeq \mathbb{E}[\delta\phi] - \gamma\,\mathbb{E}[\phi'\phi^T]\ w
\end{aligned}
$$

- The SGD iterations are:

  $\theta_{k+1} = \theta_k + \alpha_k\delta_k\phi_k \ - \alpha_k\gamma\phi'_k(\phi_k^T w_k)$

  w_k update:

  $w_{k+1} = w_k + \beta_k(\phi_k - \gamma\phi'_k)(\phi_k^T w_k)$

- $\alpha_k\delta_k\phi_k$ is the conventional TD update
- $-\alpha_k\gamma\phi'_k(\phi_k^T w_k)$ is what gives the name gradient correction. This correction makes TDC to follow the updates to minimize MSPBE objective.
- Per iteration Complexity : $\mathcal{O}(n)$

GTD2 Convergence Theorem

Consider the GTD2 algorithm with step-size sequences α_k and β_k satisfying $\beta_k = \eta\alpha_k, \eta > 0, \alpha_k, \beta_k \in (0, 1], \Sigma_{k=0}^{\infty}\alpha_k = \infty, \Sigma_{k=0}^{\infty}\alpha_k^2 < \infty$. Further, assume that (ϕ_k, r_k, ϕ'_k) is an i.i.d. sequence with uniformly bounded second moments. Let $A = \mathbb{E}[\phi_k(\phi_k - \gamma\phi'_k)^T], b = \mathbb{E}[r_k\phi_k]$, and $C = \mathbb{E}[\phi_k\phi_k^T]$. Assume that A and C are non-singular. Then the parameter vector θ_k converges with probability one to the TD fixpoint.

GTD2 Convergence Proof

Let $\rho_k^T = [w_k^T/\sqrt{\eta}, \theta_k^T]$, $g_{k+1}^T = [r_k \phi_k^T, 0^T]$.
Therefore,

$$\rho_{k+1} = \rho_k + \alpha_k \sqrt{\eta}(G_{k+1}\rho_k + g_{k+1})$$

where,

$$G_{k+1} = \begin{bmatrix} -\sqrt{\eta}\phi_k\phi_k^T & -\phi_k(\phi_k - \gamma\phi_k')^T \\ (\phi_k - \gamma\phi_k')\phi_k^T & 0 \end{bmatrix}$$

$$g_{k+1} = \begin{bmatrix} r_k\phi_k^T \\ 0 \end{bmatrix}$$

GTD2 ODE

ODE: $\dot{\rho_k} = h(\rho_k) = G\rho_k + g$ where, $G = \mathbb{E}[G_k], g = \mathbb{E}[g_k]$.
Let, $\rho_{k+1} = \rho_k + \alpha_k \sqrt{\eta}(G\rho_k + g + (G_{k+1} - G)\rho_k + (g_{k+1} - g))$,
$M_{k+1} = (G_{k+1} - G)\rho_k + (g_{k+1} - g)$, $\mathcal{F}_k = \sigma(\rho_0, M_1, .., \rho_{k-1}, M_k)$.

1. h is Lipchitz continuous with $h_\infty(\rho) = \lim_{r\to\infty} \frac{h(r\rho)}{r}$ well defined.
2. $\mathbb{E}[M_{k+1}|\mathcal{F}_k] = 0$ and $\mathbb{E}[||M_{k+1}||^2|\mathcal{F}_k] \leq c(1 + ||\rho_k||^2)$.
3. $0 < \alpha_k \leq 1$, $\Sigma_{k=1}^{\infty}\alpha_k = \infty$, $\Sigma_{k=1}^{\infty}\alpha_k^2 < \infty$.
4. $\dot{\rho} = h_\infty(\rho)$ has origin as globally asymptotically stable equilibrium.
5. $\dot{\rho} = h(\rho)$ has a globally asymptotically stable equilibrium.

The above sufficient conditions for convergence are taken from the ordinary differential equation (ODE) approach and Theorem 2.2 of Borkar and Meyn (2000).

Proof

1. h is Lipchitz continuous.

$$\begin{aligned} ||h(\rho_1) - h(\rho_2)|| &= ||G\rho_1 + g - G\rho_2 - g|| \\ &= ||G(\rho_1 - \rho_2)|| \\ &\leq ||G|| \cdot ||\rho 1 - \rho 2|| \end{aligned}$$

$h_\infty(\rho) = \lim_{r\to\infty} \frac{h(r\rho)}{r}$ is well defined.

$$\begin{aligned} \lim_{r\to\infty} \frac{h(r\rho)}{r} &= \lim_{r\to\infty} \frac{Gr\rho + g}{r} \\ &= G\rho + \lim_{r\to\infty} \frac{g}{r} \\ &= G\rho \end{aligned}$$

Proof

2.

$$
\begin{aligned}
\mathbb{E}[M_{k+1}|\mathcal{F}_k] &= \mathbb{E}[(G_{k+1} - G)\rho_k + (g_{k+1} - g)|\mathcal{F}_k] \\
&= \mathbb{E}[G_{k+1}\rho_k + g_{k+1} - (G_k\rho_k + g)|\mathcal{F}_k] \\
&= \mathbb{E}[G_{k+1}\rho_k + g_{k+1}|\mathcal{F}_k] - (G_k\rho_k + g) \\
&= (G_k\rho_k + g) - (G_k\rho_k + g) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[||M_{k+1}|||\mathcal{F}_k] = \mathbb{E}[||(G_{k+1} - G)\rho_k + (g_{k+1} - g)|||\mathcal{F}_k] \\
\text{using } \Delta\text{-inequality and cauchy-schwartz inequality} \\
\leq 2 \cdot \mathbb{E}[||(G_{k+1} - G)|| \cdot ||\rho_k|| + ||g_{k+1} - g|||\mathcal{F}_k] \\
\leq K \cdot (1 + ||\rho_k||)
\end{aligned}
$$

Proof

4. G is non-singular. Given matrix G is partitioned.

$$G = \begin{bmatrix} -\sqrt{\eta}C & -A \\ A^T & 0 \end{bmatrix}$$

$$\begin{aligned} det(G) &= det(A^T \cdot C^{-1} \cdot A) \\ &= det(C^{-1}) \cdot det(A)^2 \neq 0 \end{aligned}$$

Hence, all eigen values of G are non-zero.
The real parts of the eigen values of G are negative.
Let, $x^T = [x_1^T, x_2^T]$ such that $||x|| = 1$.

$$\begin{aligned} x^H \cdot G \cdot x &= -\sqrt{\eta} \cdot x_1^H \cdot C \cdot x_1 - x_1^H \cdot A \cdot x_2 + x_2^H \cdot A \cdot x_1 \\ &= -\sqrt{\eta} \cdot x_1^H \cdot C \cdot x_1 \end{aligned}$$

Proof

$$\therefore Re(x^H \cdot G \cdot x) = -\sqrt{\eta}||x||_C^2$$

Since, C is positive definite and all eigenvalues of G are non-zero, the real part of all eigenvalues of G are negative. Therefore, the limiting ODE

$$\rho^* = -G^{-1}g$$

is the unique asymptotically stable equilibrium.

TDC Convergence Theorem

Consider the TDC algorithm with step-size sequences α_k and β_k satisfying $\alpha_k, \beta_k \in (0,1], \Sigma_{k=0}^{\infty} \alpha_k = \infty, \Sigma_{k=0}^{\infty} \alpha_k^2 < \infty$, $\frac{\alpha_k}{\beta_k} = 0$ as k $\rightarrow \infty$. Further, assume that (ϕ_k, r_k, ϕ'_k) is an i.i.d. sequence with uniformly bounded second moments. Let $A = \mathbb{E}[\phi_k(\phi_k - \gamma\phi'_k)^T], b = \mathbb{E}[r_k \phi_k]$, and $C = \mathbb{E}[\phi_k \phi_k^T]$. Assume that A and C are non-singular. Then the parameter vector θ_k converges with probability one to the TD fixpoint.

TDC Convergence Proof

Main Idea: Beyond some integer $N_0 > 0$ updates to θ happen slowly while updates to happen w happen faster.
Intuitively from the viewpoint of slower timescale after N_0 timesteps ,w has already achieved equilibrium and from the view point of faster timescale updates to θ are quasi static

$$\theta_{k+1} = \theta_k + \beta_k \xi_k$$
$$\xi_k = \frac{\alpha_k}{\beta_k}(\delta_k \phi_k - \gamma \phi'_k {\phi_k}^{\mathsf{T}} w_k) \to 0 \text{ a.s. as } k \to \infty$$

Viewing the update equations from the faster time scale
Let $\mathcal{F}_k = \sigma(\theta_l, w_l, l \leq k; \phi_s, \phi'_s, r_s, s < k)$ be the sigma field generated by $\theta_0, w_0, \theta_{l+1}, w_{l+1}, \phi_l, \phi'_l, 0 \leq l < k$. Writing the above update equation in stochastic approximation form

$$w_{k+1} = w_k + \beta_k(\mathbb{E}[\delta_k \phi_k - \phi_k \phi_k^T w_k \mid \mathcal{F}_k] + M_{k+1})$$
$$\text{where } M_{k+1} = (\delta_k \phi_k - \phi_k \phi_k^T w_k) - (\mathbb{E}[\delta_k \phi_k - \phi_k \phi_k^T w_k \mid \mathcal{F}_k])$$

Proof

The limiting ODE for the above update equation is -

$$\dot{\theta}_k = 0, \dot{w}_k = \mathbb{E}[\delta_k \phi_k \mid \theta_k] - C w_k$$

1. $h(w_k) = \mathbb{E}[\delta_k \phi_k \mid \theta_k] - C w_k$ is Lipschitz continuous.

$$\begin{aligned} ||h(w_1) - h(w_2)|| &= ||\,\mathbb{E}[\delta_k \phi_k \mid \theta_k] - C w_1 - \mathbb{E}[\delta_k \phi_k \mid \theta_k] + C w_2|| \\ &= ||C(w_2 - w_1)|| \\ &\leq ||C|| \cdot ||w_2 - w_1|| \end{aligned}$$

$h_\infty(w) = \lim_{r\to\infty} \frac{h(rw)}{r}$ is well defined.

$$\begin{aligned} \lim_{r\to\infty} \frac{h(rw)}{r} &= \lim_{r\to\infty} \frac{\mathbb{E}[\delta_k \phi_k \mid \theta_k] - Crw}{r} \\ &= -Cw + \lim_{r\to\infty} \frac{\mathbb{E}[\delta_k \phi_k \mid \theta_k]}{r} \\ &= -Cw \end{aligned}$$

2

$$
\begin{aligned}
\mathbb{E}[M_{k+1}|\mathcal{F}_k] &= \mathbb{E}[(\delta_k\phi_k - \phi_k\phi_k^T w_k) - (\mathbb{E}[\delta_k\phi_k \mid \mathcal{F}_k] - Cw_k)|\mathcal{F}_k] \\
&= \mathbb{E}[\delta_k\phi_k - \mathbb{E}[\delta_k\phi_k \mid \mathcal{F}_k]|\mathcal{F}_k] - \mathbb{E}[\phi_k^T\phi_k w_k - Cw_k|\mathcal{F}_k] \\
&= 0 \\
||M_{k+1}|| &\leq ||\delta_k\phi_k - \mathbb{E}[\delta_k\phi_k \mid \mathcal{F}_k]|| + ||\,\mathbb{E}[\phi_k^T\phi_k w_k \mid \mathcal{F}_k] - Cw_k|| \\
&\leq K_1(1 + ||\delta_k\phi_k|| + ||Cw_k||) \\
||M_{k+1}^2|| &\leq K_2(1 + ||\delta_k\phi_k||^2 + ||Cw_k||^2) \\
\mathbb{E}[||M_{k+1}||^2|\mathcal{F}_k] &\leq K_3(1 + ||w_k||^2 + ||\theta_k^2||)
\end{aligned}
$$

The first inequality follows from application of Δ inequality, second from the fact that r_k and ϕ_k have bounded moments.
Squaring both sides, applying AM-GM inequality and taking expectations we have the last two inequalities

$$\dot{w}_k = h_\infty(w_k) = -Cw_k$$

For the above ODE, the origin is the globally asymptotically stable equilibrium since C is positive definite.

$$\dot{w}_k = \mathbb{E}[\delta_k \phi_k | \theta_k] - C w_k$$
$$w_\star = C^{-1}\, \mathbb{E}[\delta_k \phi_k | \theta_k]$$

For the above ODE, $w_\star$ is the globally asymptotically stable equilibrium since C is positive definite.

From the view point of slower timescale recursion the update equations can be written as

$$\theta_{k+1} = \theta_k + \alpha_k(\delta_k \phi_k - \gamma \phi'_k \phi_k^T C^{-1} \mathbb{E}[\delta_k \phi_k | \theta_k])$$

Let $\mathcal{G}_k = \sigma(\theta_l, l \leq k; \phi_s, \phi'_s, r_s, s < k)$ be the sigma field generated by $\theta_0, \theta_{l+1}, \phi_l, \phi'_l, 0 \leq l < k$. Writing the above update equation in stochastic approximation form

$$\begin{aligned}
\theta_{k+1} &= \theta_k + \alpha_k(\mathbb{E}[\delta_k \phi_k - \gamma \phi'_k \phi_k^T C^{-1} \mathbb{E}[\delta_k \phi_k | \theta_k] | \mathcal{G}_k] + M_{k+1}) \\
\mathrm{M}_{k+1} &= \delta_k \phi_k - \mathbb{E}[\delta_k \phi_k | \mathcal{G}_k] \\
&\quad - \gamma(\phi'_k \phi_k^T C^{-1} \mathbb{E}[\delta_k \phi_k | \theta_k] - \mathbb{E}[\phi'_k \phi_k^T C^{-1} \mathbb{E}[\delta_k \phi_k | \theta_k] | \mathcal{G}_k])
\end{aligned}$$

The limiting ODE for the above update equation is -

$$\begin{aligned}
\dot{\theta_k} &= (I - \mathbb{E}[\gamma \phi' \phi^T] C^{-1}) \mathbb{E}[\delta_k \phi_k | \theta_k] \\
&= (\mathbb{E}[\phi \phi^T] - \mathbb{E}[\gamma \phi' \phi^T]) C^{-1} \mathbb{E}[\delta_k \phi_k | \theta_k] \\
&= \mathrm{A}^T C^{-1}(-A\theta_k + b), \text{ since } \mathbb{E}[\delta_k \phi_k | \theta_k] = -A\theta_k + b
\end{aligned}$$

Proof

1. $h(\theta_k) = A^T C^{-1}(-A\theta_k + b)$ is Lipschitz continuous.

$$\begin{aligned} ||h(\theta_1) - h(\theta_2)|| &= ||A^T C^{-1}(-A\theta_1 + b - (b - A\theta_2)|| \\ &= ||A^T C^{-1} A(\theta_2 - \theta_1)|| \\ &\leq ||A^T C^{-1} A|| \cdot ||\theta_2 - \theta_1|| \end{aligned}$$

$\mathrm{h}_\infty(\theta) = \lim_{r\to\infty} \frac{h(r\theta)}{r}$ is well defined.

$$\begin{aligned} \lim_{r\to\infty} \frac{h(r\theta)}{r} &= \lim_{r\to\infty} \frac{A^T C^{-1}(b - Ar\theta)}{r} \\ &= -A^T C^{-1} A\theta \end{aligned}$$

2

$$
\begin{aligned}
\mathbb{E}[M_{k+1}|\mathcal{G}_k] &= \mathbb{E}[\delta_k\phi_k - \mathbb{E}[\delta_k\phi_k|\mathcal{G}_k]|\mathcal{G}_k] \\
&\quad - \gamma\, \mathbb{E}[(\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k]|\mathcal{G}_k] \\
&\quad + \gamma\, \mathbb{E}[\mathbb{E}[\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k]|\mathcal{G}_k])|\mathcal{G}_k]|\mathcal{G}_k] \\
&= 0 \\
||M_{k+1}|| &\leq ||\delta_k\phi_k - \mathbb{E}[\delta_k\phi_k|\mathcal{G}_k]|| \\
&\quad + \gamma||\, \mathbb{E}[(\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k] - \mathbb{E}[\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k]|\mathcal{G}_k])| \\
&\leq K_1(1 + ||\delta_k\phi_k|| + \gamma||\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k]||) \\
||M_{k+1}||^2 &\leq K_2(1 + ||\delta_k\phi_k||^2 + \gamma^2||\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k]||^2) \\
\mathbb{E}[||M_{k+1}||^2|\mathcal{F}_k] &\leq \mathbb{E}[K_2(1 + ||\delta_k\phi_k||^2 + \gamma^2||\phi'_k\phi_k^T C^{-1}\, \mathbb{E}[\delta_k\phi_k|\theta_k]||^2)] \\
&\leq K_3(1 + ||\theta_k||)^2
\end{aligned}
$$

The first inequality follows from application of Δ inequality, second from the fact that r_k and ϕ_k have bounded moments. Squaring both sides, applying AM-GM inequality and taking expectations we have the last two inequalities

$$\dot{\theta_k} = h_\infty(\theta_k) = -A^T C^{-1} A\theta_k$$

For the above ODE, the origin is the globally asymptotically stable equilibrium follows from the fact that C is positive definite and A is non singular

$$\dot{\theta_k} = A^T C^{-1}(b - A\theta)$$

$$\theta_\star = A^{-1} b$$

For the above ODE, $w_\star$ is the globally asymptotically stable equilibrium since C is positive definite and A is non singular