Proceedings of the 41st IEEE
Conference on Decision and Control
Las Vegas, Nevada USA, December 2002

Degenerate Gradient Flows: A Comparison Study of Convergence Rate Estimates [1]

S. Andersson and P. S. Krishnaprasad
Electrical and Computer Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742
{sanderss, krishna}@glue.umd.edu

Abstract

Degenerate gradient flows arise in the context of adaptive control of linear systems when the usual gradient algorithm is used for the parameter update law. It is well known that in general parameter convergence is not guaranteed without further assumptions. The standard approach utilizes the notion of a persistently exciting input and different authors have derived different convergence rate estimates. In a recent paper Brockett re-examined this issue and developed a rate estimate using a property of symmetric matrices related to the condition number. In this paper we compare two well-known convergence rate estimates from the persistently exciting point of view with Brockett's estimate through a semianalytical numerical study. We establish a common footing by relating the assumptions of each theorem to the parameters specified under the persistently exciting condition. Our analysis shows that for all parameter values Anderson's result yields a tighter bound than the other two estimates. In each case the magnitude of the difference depends on the time it takes for the uniform observability condition to hold in the persistently exciting assumption. The shorter the time is, the larger the difference is.

1 Introduction

Degenerate gradient flows are equations of the form

$$\dot{x} = -H(t,x)\frac{\partial V(x)}{\partial x} \tag{1}$$

where H is a symmetric, positive semidefinite but not positive definite matrix. Equations of this type arise when we wish to minimize a particular function but have only partial knowledge about its gradient at any given instant. Over time, however, different projections become available and it is thus possible to construct an effective descent procedure. In this paper we consider convergence rates to the zero equilibrium for degenerate flows that arise in the adaptive control of a linear system when the standard gradient algorithm is used as the parameter update law. This equation has the form

$$\dot{\phi}(t) = -w(t)w^T(t)\phi(t) \tag{2}$$

where $\phi(t)$ is the parameter error and $w(t)$ is the state of an appropriate filter. Since $w(t)w^T(t)$ is positive semidefinite it is clear that $\phi^T(t)\phi(t)$ is non-increasing but in general we cannot conclude that (2) is exponentially stable. It is well known that under an assumption of persistent excitation the equilibrium is exponentially asymptotically stable. Two convergence rate estimates, one by Sondhi and Mitra in 1976 [1] and one by Anderson in 1977 [2], are based on this assumption. A recent paper by Brockett [3] re-examined the persistently exciting hypothesis, proceeding from the notion of the *conditioning time* of the matrix $w(t)w^T(t)$ which characterizes the time interval over which the condition number (the largest eigenvalue divided by the smallest, see, e.g. [4]) of the integral of that matrix is relatively small. It is the purpose of this paper to compare the rate estimates of Anderson, Sondhi and Mitra, and Brockett. We begin in the following section by giving some useful definitions and a pair of well known lemmas that will be used in the proof of Anderson's rate estimate. In section 3 we present the three estimates we will compare. To establish the use of the persistently exciting condition we review the proof of the Anderson result but for the sake of brevity we present the other two theorems without proof, referring the reader instead to the original papers. In section 4 we turn to the comparison analysis and then conclude with a discussion of the results.

2 Background

In this section we present a few standard results for easy reference. First we need a theorem on the exponential stability of a non-autonomous system.

[1] This research was supported by ARO ODDR&E MURI97 Program Grant No. DAAG55-97-1-0114 (Center for Dynamics and Control of Smart Structures through Harvard University), ARO ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 (Center for Communicating Networked Control Systems through Boston University) and by NSF Learning and Intelligent Systems Initiative Grant CMS9720334

Theorem 2.1 (Exponential Stability) *Consider the system*

$$\dot{x} = f(t,x), \quad x \in \mathbb{R}^n \tag{3}$$

Let $x = 0$ be an equilibrium point for (3) at $t = 0$. If $\exists$ a function $v(t,x)$ and strictly positive constants $\alpha_1, \alpha_2, \alpha_3$, and δ with $\alpha_3 < \alpha_2$ such that $\forall$ x in the open ball of radius r centered at the origin for some $r > 0$ and $\forall$ $t > 0$ we have

$$\begin{array}{rcl} \alpha_1\|x\|^2 \leq & v(t,x) & \leq \alpha_2\|x\|^2 \\ \left.\frac{d}{dt}v(t,x(t))\right|_{(3)} & \leq & 0 \\ \int_t^{t+\delta} \left.\frac{d}{d\tau}v(\tau,x(\tau))\right|_{(3)} d\tau & \leq & -\alpha_3\|x(t)\|^2 \end{array}$$

then

$$\|x(t)\|^2 \leq me^{-\alpha t}\|x(0)\|^2 \tag{4}$$

where

$$m = \left[\frac{\alpha_2}{\alpha_1(1-\frac{\alpha_3}{\alpha_2})}\right] \qquad \alpha = \frac{1}{\delta}ln\left[\frac{1}{1-\frac{\alpha_3}{\alpha_2}}\right] \tag{5}$$

Proof See [5], Theorem 1.5.2. Note that there is an error in the theorem statement in that reference; specifically we additionally require $\alpha_3 < \alpha_2$. An analogous result is given as Theorem 8.5 in [6]. ■

Next we give a standard result on the uniform complete observability of a linear system under output feedback, usually known as Anderson's Lemma.

Lemma 2.2 (Anderson's Lemma) *Assume that $\forall$ $\delta > 0$ $\exists$ $k_\delta \geq 0$ such that $\forall$ $t_0 \geq 0$*

$$\int_{t_0}^{t_0+\delta} \|K(\tau)\|^2 d\tau \leq k_\delta \tag{6}$$

Let $[C,A]$ be the system

$$\begin{array}{l} \dot{x}(t) = A(t)x(t) \\ y(t) = C(t)x(t) \end{array} \tag{7}$$

and let $[C, A+KC]$ be the system with output feedback

$$\begin{array}{l} \dot{\widetilde{x}}(t) = (A(t)+K(t)C(t))\widetilde{x}(t) \\ \widetilde{y}(t) = C(t)\widetilde{x}(t) \end{array} \tag{8}$$

Let $N_1(t_0, t_0+\delta)$ and $N_2(t_0, t_0+\delta)$ be the corresponding observability grammians. That is

$$N_1(t_0,t_0+\delta) \triangleq \int_{t_0}^{t_0+\delta} \Phi_A^T(\tau,t_0)C^T(\tau)C(\tau)\Phi_A(\tau,t_0)d\tau \tag{9}$$

$$N_2(t_0,t_0+\delta) \triangleq \int_{t_0}^{t_0+\delta} \Phi_{A+KC}^T(\tau,t_0)C^T(\tau)C(\tau)\Phi_{A+KC}(\tau,t_0)d\tau \tag{10}$$

Let $\mathbb{I}$ be the identity matrix and suppose that

$$\beta_2\mathbb{I} \geq N_1(t_0,t_0+\delta) \geq \beta_1\mathbb{I} \tag{11}$$

for some constants $\beta_2 \geq \beta_1 > 0$. Then

$$\beta_2'\mathbb{I} \geq N_2(t_0,t_0+\delta) \geq \beta_1'\mathbb{I} \tag{12}$$

where

$$\beta_1' = \frac{\beta_1}{(1+\sqrt{k_\delta\beta_2})^2} \quad \beta_2' = \beta_2 e^{k_\delta\beta_2} \tag{13}$$

Proof See [5], Lemma 2.5.2. For a brief discussion and additional references see [7], Section 13.4. ■

Finally we give the definition of a persistently exciting input.

Definition 2.3 *A function $w : \mathbb{R} \to \mathbb{R}^n$ is said to be* **persistently exciting** *if $\exists$ $\alpha_1, \alpha_2, \delta > 0$ such that*

$$\alpha_2\mathbb{I} \geq \int_t^{t+\delta} w(\sigma)w^T(\sigma)d\sigma \geq \alpha_1\mathbb{I} \quad \forall t \geq 0 \tag{14}$$

3 Convergence Rate Estimates

In this section we present the three convergence rate estimates we will compare. We begin with a result based on Anderson's Lemma. The following theorem can be found in [5]. We give the proof here to illustrate the use of the persistently exciting condition.

Theorem 3.1 *Consider equation (2). If $w(t)$ is persistently exciting then*

$$\|\phi(t)\|^2 \leq me^{\alpha t}\|\phi(t_0)\|^2 \tag{15}$$

where

$$m = \frac{1}{1-\beta^2} \tag{16}$$

$$\alpha = \frac{1}{\delta}ln\left(1-\beta^2\right) \tag{17}$$

with

$$\beta^2 = \frac{\alpha_1}{(1+\sqrt{n}\alpha_2)^2} \tag{18}$$

Proof Let $v(\phi) = \frac{1}{2}\phi^T\phi$. Then along trajectories of system (2) we have

$$\dot{v} = \phi^T\dot{\phi} = -\phi^T ww^T\phi = -(w^T\phi)^2 \leq 0 \tag{19}$$

Since w is persistently exciting the system $[w^T, 0]$ is uniformly completely observable. Let $K(t) = -w$. The corresponding output feedback system is then $[w^T, -ww^T]$. Notice that

$$\int_{t_0}^{t_0+\delta} \|K(\tau)\|^2 d\tau = \int_{t_0}^{t_0+\delta} w^T(\tau)w(\tau)d\tau \tag{20}$$

$$= Tr\left(\int_{t_0}^{t_0+\delta} w(\tau)w^T(\tau)d\tau\right) \leq n\alpha_2 \tag{21}$$

where $Tr(\cdot)$ is the trace operator and n is the dimension of w. Thus by Lemma 2.2 the system $[w^T, -ww^T]$ is uniformly completely observable with constants

$$\alpha_1' = \frac{\alpha_1}{(1+\sqrt{n}\alpha_2)^2} \quad \alpha_2' = \alpha_2 e^{n\alpha_2} \tag{22}$$

So

$$\alpha_2'\|\phi(t_0)\|^2 \geq \int_{t_0}^{t_0+\delta} |w^T(\tau)\phi(\tau)|^2 d\tau \geq \alpha_1'\|\phi(t_0)\|^2 \tag{23}$$

From this we have

$$\begin{aligned}\int_{t_0}^{t_0+\delta} \dot{v}d\tau &= -\int_{t_0}^{t_0+\delta} (w^T(\tau)\phi(\tau))^2 d\tau \leq -\alpha_1'\|\phi(t_0)\|^2 \\ &= -\frac{\alpha_1}{(1+\sqrt{n}\alpha_2)^2}\|\phi(t_0)\|^2 \end{aligned}\tag{24}$$

Using the fact that $\|\phi(t)\|$ is non-increasing yields

$$\int_{t_0}^{t_0+\delta} \dot{v}d\tau \leq \frac{-\alpha_1}{(1+\sqrt{n}\alpha_2)^2}\|\phi(t)\|^2 = -\beta^2\|\phi(t)\|^2 \tag{25}$$

Then by Theorem 2.1 we have

$$\begin{aligned}\|\phi(t)\|^2 &\leq \left(\frac{1}{1-\beta^2}\right) e^{-\frac{1}{\delta}ln\left(\frac{1}{1-\beta^2}\right)t}\|\phi(t_0)\|^2 \\ &= me^{\alpha t}\|\phi(t_0)\|^2 \end{aligned}\tag{26}$$

where in the last step we used the definitions given in the statement of the theorem. ■

We turn now to a result of Sondhi and Mitra [1]

Theorem 3.2 *Consider equation (2) and assume $w(t)$ satisfies both the mixing condition*

$$\frac{1}{\delta}\int_t^{t+\delta} w(\tau)w^T(\tau)d\tau \geq \alpha_m \mathbb{I} \tag{27}$$

where $\alpha_m > 0$ and

$$\frac{1}{\delta}\int_t^{t+\delta} w^T(\tau)w(\tau)d\tau \leq L^2 \tag{28}$$

Then

$$\|\phi(t)\|^2 \leq ae^{bt}\|\phi(0)\|^2 \tag{29}$$

where

$$a = e^{-b\delta} \qquad b = max(b_1, b_2) \tag{30}$$

with

$$b_1 = \frac{1}{\delta}ln(1-s_0^2) \qquad b_2 = \frac{1}{\delta}ln(1-\rho) \tag{31}$$

where s_0 is the unique positive root of

$$(1+\delta\alpha_m + \frac{1}{2}\delta^2\alpha_m^2)^2(1-s^2) = (1+\frac{s}{4}(\delta L^2)^{\frac{5}{2}})^2 \tag{32}$$

and

$$\rho = \frac{2\alpha_m\delta}{1+L^2\delta+\frac{1}{2}L^4\delta^2} \tag{33}$$

Proof See [1], Theorem 1. ■

Finally we give the recent result of Brockett [3].

Theorem 3.3 *Consider equation (2). Let*

$$W(t) = \int_{t_0}^t w(\sigma)w^T(\sigma)d\sigma \tag{34}$$

If $\exists$ positive constants r, ϵ, and δ such that $\forall t \geq 0$ we have

$$W(t+\delta) - W(t) \geq \epsilon\mathbb{I} \tag{35}$$

and

$$Tr([W(t+\delta)-W(t)]^3) \leq r^3 \tag{36}$$

then for

$$\gamma = \sqrt{\frac{2r^3}{3(1+2\epsilon)^2}+\frac{2\epsilon}{1+2\epsilon}} - \sqrt{\frac{2r^3}{3(1+2\epsilon)^2}} \tag{37}$$

(with γ necessarily between 0 and 1) and for

$$\lambda = \frac{1}{\delta}ln(1-\gamma^2) \tag{38}$$

$\exists$ a constant d such that

$$\|\phi(t)\|^2 \leq de^{\lambda t}\|\phi(0)\|^2 \tag{39}$$

Proof See [3]. ■

4 Estimate Comparisons

4.1 Comparison of the Anderson and Brockett estimates

We compare the estimate of Theorem 3.1 to that of Brockett by first relating the assumptions used by Brockett to the persistently exciting condition. Assume that the conditions for both theorems are met. We have

$$W(t+\delta) - W(t) = \int_t^{t+\delta} w(\sigma)w^T(\sigma)d\sigma \geq \alpha_1\mathbb{I} \tag{40}$$

where the inequality comes from the persistently exciting condition. Comparing this to the assumption used by Brockett in equation (35) we take

$$\epsilon = \alpha_1 \tag{41}$$

For the next step we need the following lemma.

Lemma 4.1 *Let M be a positive semidefinite $n \times n$ matrix. Then*

$$Tr(M^3) \leq [Tr(M)]^3 \tag{42}$$

Proof Let the eigenvalues of M be $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Since M is positive semidefinite we have $\lambda_i \geq 0$ for every i. Then

$$Tr(M^3) = \sum_{i=1}^{n} \lambda_i^3 \leq \left(\sum_{i=1}^{n} \lambda_i\right)^3 = [Tr(M)]^3 \qquad (43)$$

where the first equality follows from the spectral mapping theorem and the inequality follows from the fact that the eigenvalues are nonnegative. ■

Applying Lemma 4.1 we have

$$Tr([W(t+\delta)-W(t)]^3) = Tr\left(\left[\int_t^{t+\delta} w(\sigma)w^T(\sigma)d\sigma\right]^3\right)$$
$$\leq \left(Tr\left[\int_t^{t+\delta} w(\sigma)w^T(\sigma)d\sigma\right]\right)^3 \leq n^3\alpha_2^3 \qquad (44)$$

where the inequality again follows from the persistently exciting condition and n is the dimension of w. Comparing this to the assumption used by Brockett in equation (36) gives us

$$r = n\alpha_2 \qquad (45)$$

Rewriting γ in terms of α_1, α_2, n yields

$$\gamma = \sqrt{\frac{2n^3\alpha_2^3}{3(1+2\alpha_1)^2} + \frac{2\alpha_1}{(1+2\alpha_1)}} - \sqrt{\frac{2n^3\alpha_2^3}{3(1+2\alpha_1)^2}} \qquad (46)$$

and thus

$$\gamma^2 = \frac{1}{3(1+2\alpha_1)^2}\Big[4n^3\alpha_2^3 + 6\alpha_1 + 12\alpha_1^2$$
$$-4\sqrt{n^6\alpha_2^6 + 3n^3\alpha_1\alpha_2^3 + 6n^3\alpha_1^2\alpha_2^3}\Big] \qquad (47)$$

We can now compare λ, the estimate due to Brockett, to α, the estimate due to Anderson. Starting from equation (38)

$$\lambda = \frac{1}{\delta}ln(1-\gamma^2) \qquad (48)$$
$$= \frac{1}{\delta}ln\left(\frac{1-\gamma^2}{1-\beta^2}(1-\beta^2)\right) \qquad (49)$$
$$= \frac{1}{\delta}ln(1-\beta^2) + \frac{1}{\delta}ln\left(\frac{1-\gamma^2}{1-\beta^2}\right) \qquad (50)$$
$$= \alpha + \frac{1}{\delta}ln\left(K_1(n,\alpha_1,\alpha_2)\right) \qquad (51)$$

where the last step follows from equation (17) and defines the function $K_1(n,\alpha_1,\alpha_2)$. $K_1(n,\alpha_1,\alpha_2) < 1$ would imply $\lambda < \alpha$ and thus Brockett's result would give a faster rate estimate since it is more negative. As this expression is somewhat complicated we turn to a numerical study. In Figures 1, 2, 3, and 4 we show plots of K_1 versus α_1 for different values of α_2 and for different system dimensions n. Only even values of n are considered since in the adaptive control context w is a filter vector with dimension twice that of the original state. Since $\alpha_1 \leq \alpha_2$, each curve extends only to $\alpha_1 = \alpha_2$. From the plots we see that Anderson's result gives a tighter estimate in all cases with the difference being greater for small α_2. As the dimension of the system increases the difference decreases but remains qualitatively the same. From equation (51) we see that the actual magnitude of the difference depends on δ.

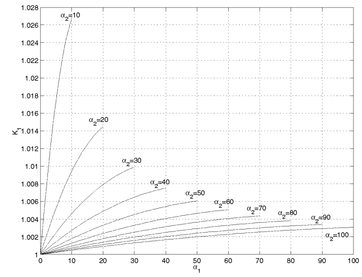

Figure 1: K_1 for $n = 2$ and select values of α_4

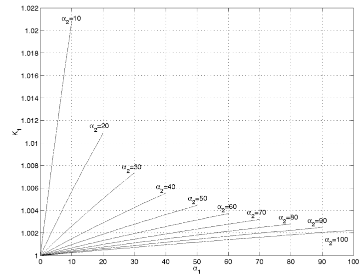

Figure 2: K_1 for $n = 4$ and select values of α_4

4.2 Comparison of the Sondhi-Mitra and Brockett estimates

To compare these two results we first express the parameters in Sondhi-Mitra's result in terms of the persistently exciting parameters α_1, α_2. Comparing the mixing condition, equation (27), to the persistently exciting condition we have

$$\alpha_m = \frac{\alpha_1}{\delta} \qquad (52)$$

Now

$$\int_t^{t+\delta} w^T(\tau)w(\tau)d\tau = Tr\left(\int_t^{t+\delta} w(\tau)w^T(\tau)d\tau\right) \leq n\alpha_2 \qquad (53)$$

with the inequality coming from the persistently exciting condition. Comparing this to equation (28) we take

$$n\alpha_2 = \delta L^2 \Rightarrow L^2 = \frac{n\alpha_2}{\delta} \qquad (54)$$

Plugging these into equation (32) we have that s_0 is the unique positive square root of

$$\left(1+\alpha_1+\frac{1}{2}\alpha_1^2\right)^2(1-s^2) = \left(1+\frac{s}{4}(n\alpha_2)^{\frac{5}{2}}\right)^2 \qquad (55)$$

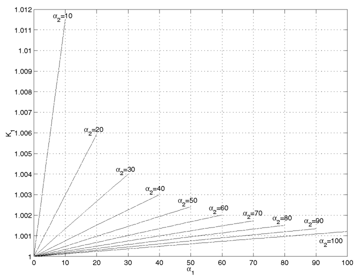

Figure 3: K_1 for $n = 8$ and select values of α_4

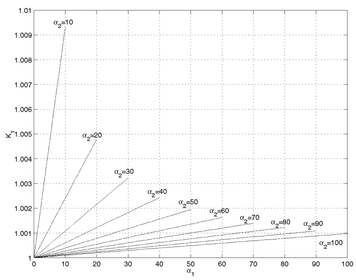

Figure 4: K_1 for $n = 10$ and select values of α_4

and into equation (33) we have

$$\rho = \frac{2\alpha_1}{1 + n\alpha_2 + \frac{1}{2}n^2\alpha_2^2} \tag{56}$$

Define

$$\xi^2 = min(s_0^2, \rho) \tag{57}$$

Using this and starting from equation (38)

$$\lambda = \frac{1}{\delta} ln(1 - \gamma^2) \tag{58}$$

$$= \frac{1}{\delta} ln \left(\frac{1 - \gamma^2}{1 - \xi^2} (1 - \xi^2) \right) \tag{59}$$

$$= \frac{1}{\delta} ln(1 - \xi^2) + \frac{1}{\delta} ln \left(\frac{1 - \gamma^2}{1 - \xi^2} \right) \tag{60}$$

$$= b + \frac{1}{\delta} ln \left(K_2(n, \alpha_1, \alpha_2) \right) \tag{61}$$

which defines the function $K_2(n, \alpha_1, \alpha_2)$. As before, if $K_2(n, \alpha_1, \alpha_2) < 1$ then $\lambda < b$ and Brockett's result gives a faster estimate than Sondhi-Mitra's. In Figures 5, 6, 7, and 8 we show plots of K_2 for the same range of parameters as we used for K_1. These plots show Brockett's result gives a tighter bound than Sondhi-Mitra's for all parameter values with the difference being greater for small α_2. As the dimension of the system grows the difference decreases. From equation (61) we see the magnitude of the difference again depends on δ and so can be quite large even if K_2 is close to one.

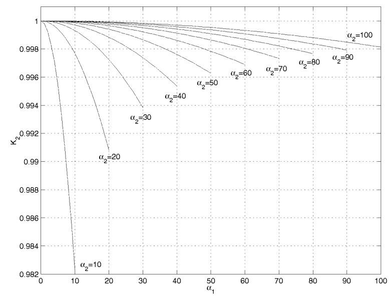

Figure 5: K_2 for $n = 2$ and select values of α_2

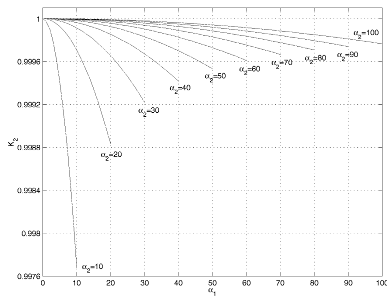

Figure 6: K_2 for $n = 4$ and select values of α_2

4.3 Comparison of the Anderson and Sondhi-Mitra results

For the sake of completeness we compare the remaining combination. Starting from equation (17) we have

$$\alpha = \frac{1}{\delta} ln(1 - \beta^2) \tag{62}$$

$$= \frac{1}{\delta} ln \left(\frac{1 - \beta^2}{1 - \xi^2} (1 - \xi^2) \right) \tag{63}$$

$$= \frac{1}{\delta} ln(1 - \xi^2) + \frac{1}{\delta} ln \left(\frac{1 - \beta^2}{1 - \xi^2} \right) \tag{64}$$

$$= b + \frac{1}{\delta} ln \left(K_3(n, \alpha_1, \alpha_2) \right) \tag{65}$$

which defines the function $K_3(n, \alpha_1, \alpha_2)$ Here, if $K_3 < 1$ Anderson's result gives a tighter estimate than Sondhi-Mitra's. In Figures 9, 10, 11, and 12 we plot K_3 over the same range of parameters as in the previous two cases. As expected we see that for all parameters the Anderson estimate gives a faster convergence rate than the Sondhi-Mitra result with the difference larger for smaller α_2 and a magnitude depending on δ.

5 Conclusions

In this paper we have presented a comparative study of three different convergence rate estimates for a degenerate gradient flow equation common in adaptive control. We considered two well-known results, one due to Anderson and one due to Sondhi and Mitra, and a

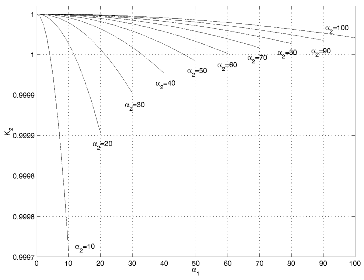

Figure 7: K_2 for $n = 8$ and select values of α_2

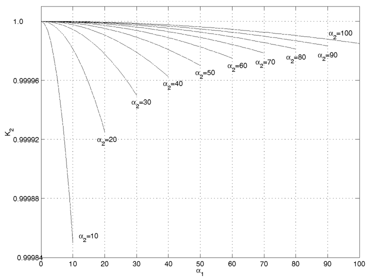

Figure 8: K_2 for $n = 10$ and select values of α_2

recent result by Brockett. Our analysis shows that Anderson's result yields a tighter estimate than the other two and that Brockett's estimate is tighter than Sondhi and Mitra's. For small δ the difference can be quite large; that is as the input becomes more strongly exciting (mixing) the Anderson result indicates a much faster rate of convergence than would be expected from either Brockett's or Sondhi and Mitra's result.

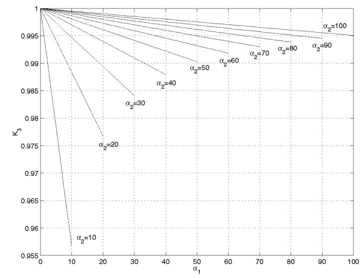

Figure 9: K_3 for $n = 2$ and select values of α_2

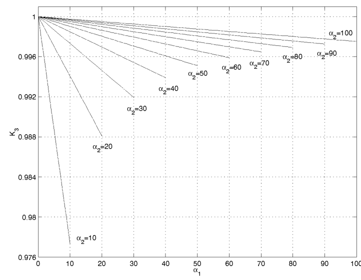

Figure 10: K_3 for $n = 4$ and select values of α_2

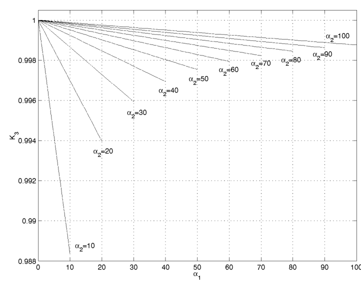

Figure 11: K_3 for $n = 8$ and select values of α_2

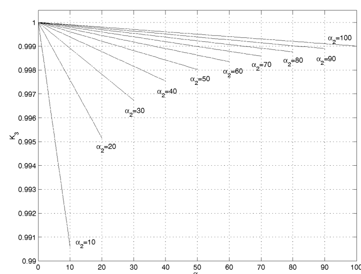

Figure 12: K_3 for $n = 10$ and select values of α_2

References

[1] M.M. Sondhi and D. Mitra. New results on the performance of a well-known class of adaptive filters. *Proc. of the IEEE*, 64(11):1583–1597, 1976.

[2] B.D.O. Anderson. Exponential stability of linear equations arising in adaptive identification. *IEEE Trans. on Automatic Control*, AC-22(1):83–88, 1977.

[3] R. Brockett. The rate of descent for degenerate gradient flows. In *Proceedings of the 2000 MTNS*, 2000.

[4] G. Strang. *Linear Algebra and Its Applications, Third Edition*. Harcourt, Brace, and Jovanovich, San Diego, 1988.

[5] S. Sastry and M. Bodson. *Adaptive Control*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[6] H. Khalil. *Nonlinear Systems, Third Edition*. Prentice Hall, 2002.

[7] H. Khalil. *Nonlinear Systems, Second Edition*. Macmillan, 1992.