The Rate of Descent for Degenerate Gradient Flows

Roger Brockett
Division of Engineering and Applied Sciences
Harvard University
33 Oxford Street
Cambridge, MA 02138, U.S.A
brockett@hrl.harvard.edu

Keywords: Gradient descent, persistent excitation, exponential stability, stability

Abstract

In this paper we investigate the question of convergence for degenerate descent procedures. The results focus attention on the role of a property of symmetric matrices analogous to, but different from, the usual condition number used in numerical linear algebra. The main result is a bound which establishes a specific rate of exponential decay for time varying linear systems with singular A matrices.

1 Introduction

The question of interest here is a particular case of that of determining the rate of convergence associated with descent algorithms of the form

$$\dot{x} = -H(t,x)\frac{\partial \phi(x)}{\partial x}$$

when $H = H^T$ is positive semidefinite but not positive definite. Such equations arise when only partial information about the gradient of the function to be minimized is available at any specific time, but over time different projections of it become available, making it possible to construct an effective descent procedure. The most frequently studied situation seems to be one that arises in adaptive control of linear systems. In this case an assumption of persistent excitation is used to prove convergence–these results are well known. Early contributions by Sondhi and MItra [1] and Anderson [2] give sufficient conditions for exponential stability. These resultls are reviewed in [3]. More recently Aeyles and Peuteman [4] have formulated a number of specific open questions in this area. In this note we examine the persistently exciting hypothesis with a view toward providing tight inequalities translating directly into explicit bounds on the a rate of convergence to be copmpared with those of [1] and [3].

The simplest problem of interest here concerns the study of the equation

$$\dot{y}(t) = -u(t)u^T(t)y(t)$$

with $u(t)$ a n-vector. For such a system $y^T(t)y(t)$ is clearly non-increasing but because the quadratic form that defines its derivative is only negative semidefinite, some further hypothesis is needed to insure exponential stability. The additional assumptions usually include the specification of a time T, we will call it the *conditioning time*, which is intended to characterize a time interval such that the condition number (the largest eigenvlaue divided by the smallest [5]) of the integral of H over that interval is relatively small. For example, if H is independent of x and if $W(t) = \int_0^t H dt$ then one often sees the hypothesis that for all $t \geq 0$ the inequalities $\eta I \geq W(t+T) - W(t) \geq \epsilon I$ are valid.

The relationship

$$\frac{d}{dt} y^T(t)y(t) = -\langle y(t), u(t)\rangle^2$$

shows that $||y(t)||$ does not decrease if $y(t)$ is orthogonal to $u(t)$. The following example describes some situations in which $||y(t)||$ decays very slowly because $y(t)$ is nearly orthogonal to $u(t)$ at all times.

Example: Let b be a constant vector and let Ω be a constant real skew-symmetric matrix. Assume that x satisfies

$$\dot{x}(t) = -e^{\Omega t} b b^T e^{-\Omega t} x(t)$$

In this case

$$W(T)) = \int_0^T e^{\Omega t} b b^T e^{-\Omega t} dt$$

and

$$x(t) = e^{\Omega t} z(t)$$

where

$$\dot{z}(t) = (-\Omega - bb^T) z$$

Because $e^{\Omega t}$ is orthogonal, the rate of decay of z is the same as that of x and is determined by the eigenvalues of $\Omega + bb^T$. In the special case

$$\Omega = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix} \; ; \; bb^T = \begin{bmatrix} m & 0 \\ 0 & 0 \end{bmatrix}$$

The eigenvalues of $-\Omega - bb^T$ are given by

$$\lambda_i = -\frac{m}{2} \pm \sqrt{\frac{m^2}{4} - \omega^2}$$

For $m^2 >> \omega^2$ we have

$$\lambda_i \approx -m + \frac{\omega^2}{m} \text{ and } -\frac{\omega^2}{m}$$

In this case it is natural to take the conditioning time to be π/ω because over the interval $[t, t+\pi/\omega]$ the matrix $W(t)$ increases by a multiple of the identity,

$$W(t+\pi/\omega) - W(t) = (m\pi/2\omega)I$$

Thus the condition number is one. Even in this ideal situation, however, the rate of decay of the least damped mode is only about

$$x(t) \approx e^{-\omega^2 t/m} x(0)$$

We call attention to the fact that for a fixed value of ω, increasing m actually slows down the rate of decay, a rather counterintuitive fact. To facilitate comparison with what follows we re-express ω^2/m in terms of the conditioning time $T = \pi/\omega$, a term that is homogeneous of degree minus one in W, and a term which is homogenous of degree zero. Let $r =(\mathrm{tr}W^3)^{1/3}$ and let T be as given. Then in terms of

$$\lambda_{max} \approx -\frac{\omega^2}{m} = \left(\frac{-2\pi^2}{4^{1/3}}\right)\frac{1}{rT}$$

we have

$$x(t) \approx e^{-\lambda_{max} t} x(0)$$

In the next section we state and prove a theorem giving a bound for a general class of systems in a form that is comparable with this computation.

2 Exponential Stability with Persistent Excitation

To begin with we establish the following lemma.

Lemma 1: Suppose that

$$\dot{y}(t) = -u(t)u^T(t)y(t)$$

Define W as

$$W(t) = \int_0^t u(\sigma)u^T(\sigma)d\sigma$$

Let $\epsilon(t)$ be the smallest eigenvalue of the symmetric, non-negative matrix $W(t)$. Then for $t > 0$ we have

$$\sqrt{\frac{y^T(0)y(0) - y^T(t)y(t)}{y^T(0)y(0)}} \geq$$

$$\sqrt{\frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon(t))^2} + \frac{2\epsilon(t)}{(1+2\epsilon(t))}} - \sqrt{\frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon(t))^2}}$$

Proof: Without loss of generality we will assume that $||y(0)|| = 1$. The general case simply involves a rescaling. We begin with the observation that

$$\frac{d}{dt}\left(y^T(t)W(t)y(t) + \frac{1}{2}y^T(t)y(t)\right) =$$

$$-2y^T(t)u(t) \cdot y^T(t)W(t)u(t)$$

Integrating this from 0 to t we see that

$$y^T(t)W(t)y(t) - \frac{1}{2}\left(1 - y^T(t)y(t)\right) =$$

$$-2\int_0^t y^T(\sigma)u(\sigma) \cdot y^T(\sigma)W(\sigma)u(\sigma)d\sigma$$

The Cauchy-Schwartz inequality implies

$$\left|\int_0^t y^T(\sigma)u(\sigma) \cdot y^T(\sigma)W(\sigma)u(\sigma)d\sigma\right| \leq$$

$$\sqrt{\int_0^t \left(y^T(\sigma)u(\sigma)\right)^2 d\sigma} \cdot \sqrt{\int_0^t \left(y^T(\sigma)W(\sigma)u(\sigma)\right)^2 d\sigma}$$

We consider the two factors on the right-hand side separately. Firtst, observe that from the differential equation for y we have

$$\frac{d}{dt}y^T(t)y(t) = -2\left(y^T(t)u(t)\right)^2$$

and so

$$\sqrt{\int_0^t \left(y^T(\sigma)u(\sigma)\right)^2 d\sigma} = \sqrt{\int_0^t -\frac{d}{2d\sigma}\left(y^T(\sigma)y(\sigma)\right)d\sigma}$$

$$= \frac{1}{\sqrt{2}}\sqrt{1 - y^T(t)y(t)}$$

We turn now to the second factor. Using the fact that $||y(0)||$ is one and $||y(t)||$ is monotone decreasing, we see that

$$\sqrt{\int_0^t \left(y^T(\sigma)W(\sigma)u(\sigma)\right)^2 d\sigma} \leq \sqrt{\int_0^t ||W(\sigma)u(\sigma)||^2 d\sigma}$$

Notice that

$$||W(t)u(t)||^2 = u^T(t)W^2(t)u(t) = \mathrm{tr}W^2(t)u^Tu(t)$$

so that

$$u^T(t)W^2(t)u(t) = \frac{1}{3}\frac{d}{dt}\mathrm{tr}W^3(t)$$

Using this in the above inequality we see that

$$\sqrt{\int_0^t \left(y^T(\sigma)W(\sigma)u(\sigma)\right)^2 d\sigma} \leq \sqrt{\frac{1}{3}\mathrm{tr}W^3(t)}$$

Putting these two inequalities together

$$y^T(t)W(t)y(t) - \frac{1}{2}\left(1 - y^T(t)y(t)\right) \leq$$

$$\sqrt{\frac{2}{3}}\sqrt{1 - y^T(t)y(t)} \cdot \sqrt{\mathrm{tr}W^3(t)}$$

We can use this to get an explicit bound on $y^T(t)y(t)$ in the following way. Let $f(t) = \sqrt{1 - y^T(t)y(t)}$, multiply the inequality by -2 and rearrange terms to get

$$f^2(t) + f(t)\frac{\sqrt{8}}{\sqrt{3}}\sqrt{\mathrm{tr}W^3(t)} \geq 2y^T(t)W(t)y(t)$$

Denoting the smallest eigenvalue of $W(t)$ by $\epsilon(t)$, we can say that

$$y^T(t)W(t)y(t) \geq \epsilon(t)y^T(t)y(t) = \epsilon(t)(1-f^2(t))$$

and thus that

$$(1+2\epsilon)f^2(t) + f(t)\sqrt{\frac{8}{3}}\sqrt{\mathrm{tr}W^3(t)} \geq 2\epsilon$$

dividing through by $1+2\epsilon$ we get

$$f^2(t) + f(t)\sqrt{\frac{8}{3}}\frac{\sqrt{\mathrm{tr}W^3(t)}}{(1+2\epsilon)} \geq \frac{2\epsilon}{(1+2\epsilon)}$$

Completing the square on the left we see that

$$\left(f(t) + \sqrt{\frac{2}{3}}\frac{\sqrt{\mathrm{tr}W^3(t)}}{(1+2\epsilon)}\right)^2 \geq \frac{2\epsilon}{(1+2\epsilon)} + \frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon)^2}$$

Taking the square root of both sides we get

$$f(t) \geq \sqrt{\frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon)^2} + \frac{2\epsilon}{(1+2\epsilon)}} - \sqrt{\frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon)^2}}$$

which establishes the lemma.

In the cases of particular interest here, ϵ is small compared wth $\sqrt{\mathrm{tr}W^3}$. Consider the Taylor series expansion

$$-a+\sqrt{a^2+b} = -a\left(1-\sqrt{1+\frac{b}{a^2}}\right) = \frac{a}{2}\frac{b}{a^2} - \frac{a}{4}\left(\frac{b}{a^2}\right)^2 + \ldots$$

Identifying b with $2\epsilon/(1+2\epsilon)$, etc. we can use this to approximate the right-hand side of the previous inequality. After squaring both sides we get

$$f^2(t) \geq k(t) \approx \frac{3\epsilon^2(t)}{2\mathrm{tr}W^3(t)}$$

This, in turn, can be factored as the product of a term that is homogeneous of degree zero in $W(t)$ and a term which is homogeneous of degree minus one in $W(t)$,

$$f^2(t) = \frac{3\epsilon^2(t)}{2(\mathrm{tr}W^3(t))^{2/3}}\frac{1}{(\mathrm{tr}W^3(t))^{1/3}}$$

We can interpret the first factor as being analogous to the square of the reciprocal of the condition number of $W(t)$. Specifically, it is the square of the smallest eigenvalue of $W(t)$ divided by the sum of the cubes of all the eigenvalues of $W(t)$, raised to the power 2/3. The second factor is the reciprocal of a measure of the size of $W(t)$. Taking into account the fact that f measures the decay over a period, if we take that period to be T then

$$\lambda_{max} \approx -\left(\frac{6\epsilon^2(T)}{(\mathrm{tr}W^3T(T))^{2/3}}\right)\left(\frac{1}{(\mathrm{tr}W(T))^{1/3}T}\right)$$

The second factor can be identified with the factor $1/rT$ appearing in the example in the introduction.

The lemma just proven provides the basic inequality required to prove the following result.

Theorem: Let y satisfy the equation

$$\dot{y}(t) = -u(t)u^T(t)y(t)$$

Define W as

$$W(t) = \int_0^t u(\sigma)u^T(\sigma)d\sigma$$

Assume that there exist constants ϵ, r and T such that for all $t \geq 0$

$$W(t+T) - W(t) \geq \epsilon \cdot I$$

and

$$\mathrm{tr}(W(t+T)-W(t))^3 \leq r^3$$

Then for

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

(necessarily between zero and one) and for

$$\lambda = \frac{1}{T}\ln(1-\gamma^2)$$

there is a constant d such that

$$||y(t)||^2 \leq d||y(0)||^2 e^{\lambda t}$$

Proof: From the lemma we see that

$$\frac{y^T(0)y(0) - y^T(T)y(T)}{y^T(0)y(0)} \geq \gamma^2$$

which implies

$$y^T(T)y(T) \leq (1-\gamma^2)y^T(0)y(0)$$

Thus over an interval of length T, the square of the norm $||y(t)||^2$ shrinks by the factor $1-\gamma^2$. Solving

$$e^{\lambda T} = 1-\gamma^2$$

we get $\lambda = \ln(1-\gamma^2)/T$.

3 A Generalization

Consider now the more general situation

$$\dot{y}(t) = -U(t)U^T(t)y(t)$$

with $U(t)$ an n by p matrix. In this case we have

$$\frac{d}{dt}y^T(t)y(t) = -2\langle U(t)y(t), U(t)y(t)\rangle$$

As above,

$$\frac{d}{dt}\left(y^T(t)W(t)y(t) + \frac{1}{2}y^T(t)y(t)\right) =$$

$$-2y^T(t)(U^T(t)U(t)W(t)+W(t)U(t)U^T(t))y(t)$$

Because we still have

$$\frac{d}{dt}\mathrm{tr}W^3(t)=\mathrm{tr}W^2(t)U(t)U^T(t)+$$

$$\mathrm{tr}W(t)U(t)U^T(t)W(t)+\mathrm{tr}(t)U(t)U^T(t)W^2$$

the above argument goes through without change. We state without further comment the following lemma.

Lemma 2: Suppose that

$$\dot{y}(t)=-U(t)U^T(t)y(t)$$

Define W as

$$W(t)=\int_0^t U(\sigma)U^T(\sigma)d\sigma$$

Let $\epsilon(t)$ be the smallest eigenvalue of the symmetric, non-negative matrix $W(t)$. Then for $t>0$ we have

$$\sqrt{\frac{y^T(0)y(0)-y^T(t)y(t)}{y^T(0)y(0)}}\geq$$

$$\sqrt{\frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon(t))^2}+\frac{2\epsilon(t)}{(1+2\epsilon(t))}}-\sqrt{\frac{2\mathrm{tr}W^3(t)}{3(1+2\epsilon(t))^2}}$$

Acknowledgements

This work was funded by the Army grant DAAG 55 97 1 0114, Brown Univ. Army DAAH 04-96-1-0445, and MIT Army DAAL03-92-G-0115.

References

1. M. M. Sondhi and D. Mitra, *New Results on the Performance of a Well-Known Class of Adaptive Filters*, Proceedings of the IEEE, vol. 64, pp. 1583-1597, 1976.

2. Anderson B.D.O, "*Exponential stability of linear Equations arising in adaptive control*," IEEE Trans. Automatic Control, vol. 22, pp. 84-88, (1977).

3. S. Sastry and M. Bodson, *Adaptive Control*, Prentice Hall, Englewood Cliffs, New Jersey, 1989.

4. Aeyels D. and Peuteman J. ,"*Uniform asymptotic stability of linear time-varying systems*," in *Open Problems in Mathematical Systems and Control Theory* (V. Blondel et al. Eds.) Springer-Verlag, London, 1998.

5. Strang, Gilbert, *Linear Algebra and its Applications*, Harcort, Brace and Jovanovich, San Diego, 1988.