11-1-1991

Minimizing Quotient Space Norms Using Penalty Functions

Stefen Hui
San Diego State University San Diego, CA

Walter E. Lillo
Purdue University

Stanislaw H. Zak
Purdue University

MINIMIZING QUOTIENT SPACE NORMS USING PENALTY FUNCTIONS

Stefen Hui
Department of Mathematical Sciences
San Diego State University
San Diego, CA 92182

Walter E. Lillo
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

Stanislaw H. Żak
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

Abstract

A penalty function method approach is proposed to solve the general problem of quotient space norms minimization. A new class of penalty functions is introduced which allows one to transform constrained optimization problems of quotient space norms minimization by unconstrained optimization problems. The sharp bound on the weight parameter is given for which constrained and unconstrained problems are equivalent. Also a computationally efficient bound on the weight parameter is given. Numerical examples and computer simulations illustrate the results obtained.

Key Words:

Minimum norm problems, Constrained optimization, Penalty method, Linear equations.

AMS Subject Classification:

15A60, 15A06, 34A30, 90C30

Abbreviated Title:

Minimizing Quotient Space Norms

I. INTRODUCTION

In this paper we propose a novel approach to solving a system of consistent linear equations $Ax=b$ while minimizing $||x||_p$. Here $||\cdot||_p$ stands for the ℓ^p-norm of $\mathbb{R}^n$ defined by

$$||x||_p = \left[\sum_{j=1}^{n} |x_j|^p\right]^{\frac{1}{p}} \quad \text{for } 1 \leq p < \infty ,$$

and

$$||x||_\infty = \max_{1 \leq j \leq n} |x_j|$$

The above problem is also known in the literature as the minimum norm or the quotient space norm problem (see e.g. Luenberger [6]).

Problems of this form arise in various disciplines. Here we give as an example a problem in the control of discrete dynamic systems. Consider a dynamic system modeled by the difference equation

$$\xi_{k+1} = F\,\xi_k + G\,u_k ,$$

where $\xi_k \in \mathbb{R}^m$, $u_k \in \mathbb{R}^t$, $F \in \mathbb{R}^{m \times m}$, $G \in \mathbb{R}^{m \times t}$ $(m > t)$.

If we iteratively apply the previous equation, we obtain the following

$$\xi_N = F^N\,\xi_0 + F^{N-1}\,Gu_0 + \ldots + F\,G\,u_{N-2} + G\,u_{N-1} .$$

We assume that our system model is completely controllable. This implies that we can drive the system trajectory to an arbitrary desired state ξ_d regardless of the initial state ξ_0 (see e.g. [3] for more details). Thus for sufficiently large N, $(N \geq m)$ we can find a sequence of inputs $(u_0, u_1, \ldots, u_{N-1})$ such that $\xi_d = \xi_N$.

Let

$$A = [G, FG, \ldots F^{N-1}G] , \quad b = \xi_d - F^N \xi_0 , \quad z = [u_{N-1}, \ldots, u_1, u_0]^T .$$

Then the problem of finding a control sequence transferring the system from the initial state ξ_0 to the desired state ξ_d is reduced to solving the system of linear equations $Az = b$. Note that the matrix A is of full rank since we have assumed that the pair (F,G) is controllable and $N \geq m$. From the above it is clear that for $t > 1$ and $N \geq m$ there is an infinite number of feasible solutions to the problem. Therefore, secondary criteria are often imposed on the control law. For example, one may want to find an optimal control sequence using the following criteria

$$\text{minimize } ||z||_p \, , \quad 1 \leq p \leq \infty$$

subject to

$$Az = b \, .$$

The solution corresponding to $p=1$ is often referred to as the minimum fuel, p=2 - minimum energy, and $p=\infty$, minimum amplitude. Because of the importance of these problems they have been studied extensively (see eg [1], [2], [4], [5]). The cases of $p=1$ and $p=\infty$ are somewhat more complex. There are some algorithms based on results from linear programming ([1], [2]) and iterative procedures based on the steepest descent method for constrained optimization problems ([4]) which have been proposed to solve these problems. The above proposed algorithms however, are not computationally efficient. In applications such as real time control the speed at which a solution can be obtained is of the utmost importance. It is for this reason that we propose a new approach to the problem of quotient space norms minimization.

In this paper we propose to use the penalty function method. We propose a new class of penalty functions which allows one to transform constrained optimization problems of quotient space norms minimization into equivalent unconstrained minimization problems.

II. PRELIMINARIES

Recall that we are concerned with the problem

$$\text{minimize } ||z||_p , \quad 1 \leq p \leq \infty ,$$

subject to

$$Az = b ,$$

where

$$A \in \mathbb{R}^{m \times n} , \quad m < n , \quad \text{rank } A = m .$$

Any $z \in \mathbb{R}^n$ which minimizes the objective function over the given constraint set is called a minimizer. We call two minimization problems equivalent if they have the same minimums and minimizers. In this paper, we show that the constrained problem:

$$\min ||z||_p \text{ subject to } Az = b$$

is equivalent to the unconstrained problem

$$\min E_{p,s,c}(z) = \min (||z||_p + c||Az - b||_s)$$

for $1 \leq s \leq \infty$ and c greater than a constant c_0 which depends on p, s, and A. The constant c_0 will be given explicitly in Theorem 1.

The technique of transforming a constrained problem into an unconstrained problem by adding a function of the constraint, called the penalty function, to the objective function is known as the Penalty Function Method (see Luenberger [7]). In general, the constrained problem and penalized unconstrained problem are not equivalent but are related by the fact that as $c \rightarrow \infty$, the limit points of the minimizers of the unconstrained problem are minimizers of the constrained problem. We first introduce a simple change of variables which makes the development and the proof of Theorem 1 notationally easier. Let $r \in \mathbb{R}^n$ be such that $Ar = -b$ and let $x = z + r$. Note that one such r is $r = -A^T(AA^T)^{-1}b$. Then the constrained problem has the form

$$\min \|x - r\|_p \quad \text{subject to } Ax = 0$$

and the unconstrained problem has the form

$$\min \left(\|x - r\|_p + c\|Ax\|_s\right) .$$

Clearly this change of variables does not affect the problem. Let

$$E_{p,s,c}(x) = \|x - r\|_p + c\|Ax\|_s .$$

When the constrained and unconstrained problems are equivalent for sufficiently large c, the penalty function is called an exact penalty function. (See Luenberger [7]).

III. MAIN RESULTS

Before stating our main result precisely, we introduce some notations. For $1 \leq p \leq \infty$, we denote its conjugate exponent by p', that is, $\frac{1}{p} + \frac{1}{p'} = 1$. We adopt the convention that $1' = \infty$ and $\infty' = 1$. For any matrix $M \in \mathbb{R}^{k \times \ell}$, $1 \leq k, \ell < \infty$, we define for $1 \leq p, s \leq \infty$:

$$\|M\|_{p,s} = \sup_{x \in \mathbb{R}^\ell, x \neq 0} \frac{\|Mx\|_p}{\|x\|_s} .$$

Thus, $\|M\|_{p,s}$ is the operator norm of

$$M : (\mathbb{R}^\ell, \|\cdot\|_s) \rightarrow (\mathbb{R}^k, \|\cdot\|_p) .$$

When p=s, we just write $\|M\|_p$.

We can now state our main result.

Theorem 1. Let $1 \leq p,s \leq \infty$. Then for $c > \|A^T(AA^T)^{-1}\|_{p,s}$, the constrained problem

$$\min \|x - r\|_p \text{ subject to } Ax = 0$$

is equivalent to the unconstrained problem

$$\min E_{p,s,c}(x) = \min \left(||x-r||_p + c||Ax||_s\right) .$$

Furthermore, the bound is sharp in the sense that for each p,s, $1 \leq p,s \leq \infty$, there exist A,b so that the constrained and unconstrained problems are not equivalent if $c < ||A^T(AA^T)^{-1}||_{p,s}$.

The proof of Theorem 1 will be given in this section after some preliminary results. We will also give an upper bound of $||A^T(AA^T)^{-1}||_{p,s}$ which may be useful in practice.

From Theorem 1, we see that our original constrained problem can be transformed into an unconstrained problem which gives the same solutions. Since $E_{p,s,c}(x)$ is convex for $1 \leq p,s \leq \infty$ (strictly convex for $1 \leq p,s < \infty$), the unconstrained problem can be solved by a number of well-known methods. In this paper we use a continuous gradient descent method for the solution of the unconstrained problem. The trajectories of the continuous gradient descent are governed by:

$$\frac{dx}{dt} = - \nabla E_{p,s,c}(x) .$$

Note that even though $\nabla E_{p,s,c}(x)$ may not exist everywhere, the set where it is not defined has n-dimensional Lebesgue measure zero. We would like to converge to a minimizer of $||x-r||_p$ subject to $Ax=0$ starting from an arbitrary $x \in \mathbb{R}^n$. Note that any $x \in \mathbb{R}^n$ can be orthogonally decomposed as

$$x = (I_n - A^T(AA^T)^{-1}A)x + A^T(AA^T)^{-1}Ax .$$

Thus the Euclidean distance from x to $\{x \mid Ax = 0\}$ is $||A^T(AA^T)^{-1}Ax||_2$. Therefore, for the trajectories of $\frac{dx}{dt} = - \nabla E_{p,s,c}(x)$ to converge to the feasible set $\{x \mid Ax=0\}$ it is sufficient that

$$\frac{d}{dt}\left(\frac{1}{2}\,||A^T(AA^T)^{-1}Ax||_2^2\right)$$

$$= x^T A^T (AA^T)^{-1} A\dot{x}$$

$$= -x^T A^T (AA^T)^{-1} A \,\nabla E_{p,s,c} < 0 \qquad \forall\ x \notin \{x\,|\,Ax=0\}\,.$$

If the above condition is satisfied then a trajectory is guaranteed to hit the feasible region and stay there thereafter searching for a minimizer of $||x-r||_p$. It turns out that this is the case when $c > ||A^T(AA^T)^{-1}||_{p,s}$ as will become evident from the proof of Theorem 1. In the proof of Theorem 1 we need the following technical lemma. The proof is included for the convenience of the reader.

Lemma 1. Let $v \in \mathbb{R}^\ell$ and $M \in \mathbb{R}^{k\times\ell}$. Suppose $1 \le p,s \le \infty$.

(a) If $s \le p$, then

$$||v||_p \le ||v||_s \le \ell^{\frac{1}{s}-\frac{1}{p}}\,||v||_p\,.$$

(b) For $1 \le p_0,\ s_0 \le \infty$, we have

$$\lim_{p\to p_0} ||v||_p = ||v||_{p_0}$$

and

$$\lim_{\substack{p\to p_0\\ s\to s_0}} ||M||_{p,s} = ||M||_{p_0,s_0}.$$

Proof. (a) It is elementary that if $\beta \ge 1$ and $a,b \ge 0$, then $(a+b)^\beta \ge a^\beta + b^\beta$. By induction we have

$$||v||_p = \left[\sum_{j=1}^{\ell} |v_j|^p\right]^{\frac{1}{p}}$$

$$= \left[\sum_{j=1}^{\ell} (|v_j|^s)^{\frac{p}{s}}\right]^{\frac{1}{p}}$$

$$\leq \left[\sum_{j=1}^{\ell} |v_j|^s\right]^{\frac{p}{s}\frac{1}{p}}$$

$$= ||v||_s \, .$$

For the second inequality, we use Hölder's inequality ([8] p. 21):

$$||v||_s = \left[\sum_{j=1}^{\ell} |v_j|^s\right]^{\frac{1}{s}}$$

$$\leq \left\{\left[\sum_{j=1}^{\ell} |v_j|^{s\frac{p}{s}}\right]^{\frac{s}{p}} \left[\sum_{j=1}^{\ell} 1\right]^{1-\frac{s}{p}}\right\}^{\frac{1}{s}}$$

$$= \ell^{\frac{1}{s}-\frac{1}{p}} ||v||_p \, .$$

(b) The first equality involving v follows from the continuity of the exponential function.

To verify the second equality, observe that by compactness, there exists $x^* \in \mathbb{R}^\ell$ so that $||x^*||_{s_0} = 1$ and $||M||_{p_0,s_0} = ||Mx^*||_{p_0}$. Then by part (a) and the definition of $||M||_{p,s}$,

$$\begin{aligned}||M||_{p_o,s_o} &= ||Mx^*||_{p_o} \\ &\le k^{\left|\frac{1}{p_o}-\frac{1}{p}\right|} ||Mx^*||_p \\ &\le k^{\left|\frac{1}{p_o}-\frac{1}{p}\right|} ||M||_{p,s}\, ||x^*||_s \\ &\le k^{\left|\frac{1}{p_o}-\frac{1}{p}\right|} ||M||_{p,s}\, \ell^{\left|\frac{1}{s_o}-\frac{1}{s}\right|} ||x^*||_{s_o} \\ &= k^{\left|\frac{1}{p_o}-\frac{1}{p}\right|} \ell^{\left|\frac{1}{s_o}-\frac{1}{s}\right|} ||M||_{p,s} .\end{aligned}$$

By symmetry, we also have

$$||M||_{p,s} \le k^{\left|\frac{1}{p_o}-\frac{1}{p}\right|} \ell^{\left|\frac{1}{s_o}-\frac{1}{s}\right|} ||M||_{p_o,s_o} .$$

It is immediate then

$$\lim_{\substack{p\to p_o \\ s\to s_o}} ||M||_{p,s} = ||M||_{p_o,s_o} .$$

□

Proof of Theorem 1. For $1 \le p,\ s \le \infty$ and $c > 0$, let

$$E_{p,s,c}(x) = ||x-r||_p + c||Ax||_s .$$

We first study the case of $1 < p,\ s < \infty$. In this case, $\nabla E_{p,s,c}$ exists except at $x=r$ and $Ax=0$. For $j=1,\ldots,m$, let a_j be the j-th row of A. Where it is defined,

$$\nabla E_{p,s,c}(x) = \frac{1}{||x-r||_p^{p-1}} \begin{bmatrix} \mathrm{sgn}(x_1-r_1)\,|\,x_1-r_1\,|^{p-1} \\ \vdots \\ \mathrm{sgn}(x_n-r_n)\,|\,x_n-r_n\,|^{p-1} \end{bmatrix} + \frac{cA^T}{||Ax||_s^{s-1}} \begin{bmatrix} \mathrm{sgn}(a_1x)\,|\,a_1x\,|^{s-1} \\ \vdots \\ \mathrm{sgn}(a_mx)\,|\,a_mx\,|^{s-1} \end{bmatrix} .$$

We will now show that $\nabla E_{p,s,c}(x) \neq 0$ for $c > ||A^T(AA^T)^{-1}||_{p,s}$. Consider the second term in the product $[x^TA^T(AA^T)^{-1}A]\nabla E_{p,s,c}(x)$:

$$\mathrm{II} = \frac{c\,x^T A^T (AA^T)^{-1} AA^T}{||Ax||_s^{s-1}} \begin{bmatrix} \mathrm{sgn}(a_1 x)\,|\,a_1 x\,|^{s-1} \\ \vdots \\ \mathrm{sgn}(a_m x)\,|\,a_m x\,|^{s-1} \end{bmatrix}$$

$$= \frac{c\,x^T A^T}{||Ax||_s^{s-1}} \begin{bmatrix} \mathrm{sgn}(a_1 x)\,|\,a_1 x\,|^{s-1} \\ \vdots \\ \mathrm{sgn}(a_m x)\,|\,a_m x\,|^{s-1} \end{bmatrix}$$

$$= \frac{c}{||Ax||_s^{s-1}} \sum_{j=1}^{m} |\,a_j x\,|^s$$

$$= c\,||Ax||_s \,.$$

We now estimate the first term in the product $[x^T A^T (AA^T)^{-1} A]\,\nabla E_{p,s,c}(x)$. By Hölder's inequality:

$$|\,\mathrm{I}\,| = \left| \frac{x^T A^T (AA^T)^{-1} A}{||x-r||_p^{p-1}} \begin{bmatrix} \mathrm{sgn}(x_1 - r_1)\,|\,x_1 - r_1\,|^{p-1} \\ \vdots \\ \mathrm{sgn}(x_n - r_n)\,|\,x_n - r_n\,|^{p-1} \end{bmatrix} \right|$$

$$\leq \frac{||x^T A^T (AA^T)^{-1} A||_p}{||x-r||_p^{p-1}} \left|\left| \begin{bmatrix} \mathrm{sgn}(x_1 - r_1)\,|\,x_1 - r_1\,|^{p-1} \\ \vdots \\ \mathrm{sgn}(x_n - r_n)\,|\,x_n - r_n\,|^{p-1} \end{bmatrix} \right|\right|_{p'} .$$

Observe that

$$\left|\left| \begin{bmatrix} \mathrm{sgn}(x_1 - r_1)\,|\,x_1 - r_1\,|^{p-1} \\ \vdots \\ \mathrm{sgn}(x_n - r_n)\,|\,x_n - r_n\,|^{p-1} \end{bmatrix} \right|\right|_{p'}$$

$$= \left[\sum_{j=1}^{n} |x_j - r_j|^{(p-1)p'} \right]^{\frac{1}{p'}}$$

$$= \left[\sum_{j=1}^{n} |x_j - r_j|^{p} \right]^{\frac{p-1}{p}}$$

$$= ||x-r||_p^{p-1}$$

since $\frac{1}{p} + \frac{1}{p'} = 1$.

Thus

$$|I| \leq ||x^T A^T (AA^T)^{-1} A||_p$$

$$= ||A^T (AA^T)^{-1} Ax||_p$$

$$\leq ||A^T (AA^T)^{-1}||_{p,s} \; ||Ax||_s$$

by the definition of $||A^T(AA^T)^{-1}||_{p,s}$. Therefore when $x \neq r$ and $Ax \neq 0$, $\nabla E_{p,s,c}$ is defined and

$$[x^T A^T (AA^T)^{-1} A] \nabla E_{p,s,c}(x)$$

$$= I + II$$

$$\geq c||Ax||_s - ||A^T (AA^T)^{-1}||_{p,s} \; ||Ax||_s$$

$$= (c - ||A^T (AA^T)^{-1}||_{p,s}) ||Ax||_s \, .$$

We conclude that for $x \neq r$, $Ax \neq 0$, and $c > ||A^T(AA^T)^{-1}||_{p,s}$, the gradient $\nabla E_{p,s,c}(x)$ is defined and is nonzero. Since

$$\lim_{||x|| \to \infty} E_{p,s,c}(x) = \infty ,$$

the minimum of $E_{p,s,c}(x)$ must be achieved at $x = r$ or $Ax = 0$. Let $\tilde{r} = (I - A^T(AA^T)^{-1}A)r$. Then $\tilde{r} \in \{Ax=0\}$ and $E_{p,s,c}(\tilde{r}) = ||A^T(AA^T)^{-1}A\,r||_p$. If

$c > ||A^T(AA^T)^{-1}||_{p,s}$, then

$$\begin{aligned} E_{p,s,c}(r) &= c||Ar||_s \\ &> ||A^T(AA^T)^{-1}Ar||_p \\ &= E_{p,s,c}(\tilde{r}) \,. \end{aligned}$$

Thus the minimum is not achieved at $x=r$. The only remaining possibility is that the minimum is achieved on $\{Ax=0\}$. The proof for the case $1 < p,s < \infty$ is complete.

The technique we used above does not apply directly when at least one of p and s is 1 or ∞. The main difficulty being that $x=r$ is no longer the only critical point off $\{Ax=0\}$. We use what we have proved above and a limiting argument to avoid this difficulty.

We will prove the case of $p=1$ and $s=\infty$. The other cases are very similar. For $1 \le p,s \le \infty$, let $m_{p,s,c}$ be minimizers of $E_{p,s,c}(x)$. Suppose $c > ||A^T(AA^T)^{-1}||_{1,\infty}$. Then by Lemma 1, there exist $\alpha,\beta > 0$ so that for $1 \le p < 1+\alpha$ and $s > \beta$, we have $c > ||A^T(AA^T)^{-1}||_{p,s}$. Let us now assume $1 < p < 1+\alpha$ and $\infty > s > \beta$. Then $m_{p,s,c} \in \{Ax=0\}$ and is independent of c and s. We can therefore write $m_p = m_{p,s,c}$. By Lemma 1 and the fact that m_p is a minimizer of $E_{p,s,c}(x)$, we have

$$\begin{aligned} ||m_p - r||_p &= E_{p,s,c}(m_p) \\ &\le E_{p,s,c}(m_{1,\infty,c}) \\ &= ||m_{1,\infty,c} - r||_p + c||Am_{1,\infty,c}||_s \\ &\le ||m_{1,\infty,c} - r||_1 + c\, m^{\frac{1}{s}} \, ||Am_{1,\infty,c}||_\infty \\ &\le m^{\frac{1}{s}} \, E_{1,\infty,c}\,(m_{1,\infty,c}) \end{aligned}$$

$$\leq m^{\frac{1}{s}} E_{1,\infty,c}(m_p)$$

$$= m^{\frac{1}{s}} ||m_p - r||_1$$

$$\leq m^{\frac{1}{s}} n^{1-\frac{1}{p}} ||m_p - r||_p .$$

Since m_p is independent of s, we can let $s \rightarrow \infty$ and then $p \rightarrow 1$ to see that $\lim_{p \rightarrow 1} ||m_p - r||_p$ exists and

$$\lim_{p \rightarrow 1} ||m_p - r||_p = ||m_{1,\infty,c} - r||_1 + c||A m_{1,\infty,c}||_\infty .$$

We conclude that $||m_{1,\infty,c} - r||_1 + c||A m_{1,\infty,c}||_\infty$ is independent of $c > ||A^T(AA^T)^{-1}||_{1,\infty}$. To see that this forces $m_{1,\infty,c} \in \{Ax=0\}$, choose $\tilde{c}$ so that $c > \tilde{c} > ||A^T(AA^T)^{-1}||_{1,\infty}$. If $A m_{1,\infty,c} \neq 0$, then

$$E_{1,\infty,\tilde{c}}(m_{1,\infty,c}) = ||m_{1,\infty,c} - r||_1 + \tilde{c}||A m_{1,\infty,c}||_\infty$$

$$< ||m_{1,\infty,c} - r||_1 + c||A m_{1,\infty,c}||_\infty$$

$$= ||m_{1,\infty,\tilde{c}} - r||_1 + \tilde{c}||A m_{1,\infty,\tilde{c}}||_\infty$$

$$= E_{1,\infty,\tilde{c}}(m_{1,\infty,\tilde{c}}) .$$

This contradicts the fact that $m_{1,\infty,\tilde{c}}$ is a minimizer of $E_{1,\infty,\tilde{c}}(x)$. Thus we must have $A m_{1,\infty,c} = 0$ for $c > ||A^T(AA^T)^{-1}||_{1,\infty}$. The proof for the case p=1 and $s=\infty$ is complete. The proofs of the other cases are similar and will be omitted.

We now show that for any $1 \leq p,s \leq \infty$, there exist A and r so that the minimizers of $E_{p,s,c}(x)$ are not on $\{Ax=0\}$ for $c < ||A^T(AA^T)^{-1}||_{p,s}$. Suppose $1 \leq p,s \leq \infty$. Let $A = [0 \ \ 1]$ and $r = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then

$$E_{p,s,c}(x) = (|x_1|^p + |x_2 - 1|^p)^{\frac{1}{p}} + c|x_2| .$$

Clearly the minimum of $E_{p,s,c}(x)$ on $\{Ax=0\}$ is 1 and is achieved at $\begin{bmatrix}0\\0\end{bmatrix}$. We have

$$||A^T(AA^T)^{-1}||_{p,s} = || \begin{bmatrix}0\\1\end{bmatrix} ||_{p,s} = 1 .$$

For $c < ||A^T(AA^T)^{-1}||_{p,s} = 1$, we have

$$E_{p,s,c}\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = c < 1 .$$

Thus the minimizers of $E_{p,s,c}(x)$ are not on $\{Ax=0\}$. The proof of Theorem 1 is now complete. □

Remark. The technique used in the proof of the case $p=1$ and $s=\infty$ can be used to show the continuity of the minimums and the minimizers for $1 < p,s < \infty$.

We next give a bound for $||A^T(AA^T)^{-1}||_{p,s}$ which involves only the dimensions of A and the least eigenvalue of AA^T. The least eigenvalue of AA^T may be easier to compute them $||A^T(AA^T)^{-1}||_{p,s}$ which involves an inverse and an operator norm.

Proposition 1. Suppose $A \in \mathbb{R}^{m\times n}$, $m < n$, is a full rank matrix. Let $1 \le p,s \le \infty$. Then

$$||A^T(AA^T)^{-1}||_{p,s} \le \frac{K}{\sqrt{\lambda_{min}(AA^T)}} ,$$

where

$$K = \begin{cases} m^{\frac{1}{2}-\frac{1}{s}} & \text{for} \quad p \geq 2, \; s \geq 2 \\ 1 & \text{for} \quad p \geq 2, \; s < 2 \\ n^{\frac{1}{p}-\frac{1}{2}} m^{\frac{1}{2}-\frac{1}{s}} & \text{for} \quad p < 2, \; s \geq 2 \\ n^{\frac{1}{p}-\frac{1}{2}} & \text{for} \quad p < 2, \; s < 2 \end{cases}.$$

Here $\lambda_{\min}(AA^T)$ denotes the least eigenvalue of the positive definite matrix AA^T.

Proof. Suppose $1 \leq p,s < 2$. By compactness, there exists $z \in \mathbb{R}^m$ so that $||z||_s = 1$ and $||A^T(AA^T)^{-1}||_{p,s} = ||A^T(AA^T)^{-1}z||_p$. By Lemma 1,

$$\begin{aligned} ||A^T(AA^T)^{-1}||_{p,s} &= ||A^T(AA^T)^{-1}z||_p \\ &\leq n^{\frac{1}{p}-\frac{1}{2}} ||A^T(AA^T)^{-1} z||_2 \\ &= n^{\frac{1}{p}-\frac{1}{2}} [z^T(AA^T)^{-1}AA^T(AA^T)^{-1} z]^{\frac{1}{2}} \\ &= n^{\frac{1}{p}-\frac{1}{2}} [z^T(AA^T)^{-1}z]^{\frac{1}{2}} \\ &\leq n^{\frac{1}{p}-\frac{1}{2}} \sqrt{\lambda_{\max}(AA^T)^{-1}} \; ||z||_2 \\ &\leq \frac{n^{\frac{1}{p}-\frac{1}{2}} ||z||_s}{\sqrt{\lambda_{\min}(AA^T)}} \\ &= \frac{n^{\frac{1}{p}-\frac{1}{2}}}{\sqrt{\lambda_{\min}(AA^T)}} . \end{aligned}$$

The proofs for the other cases are similar with the only difference being how Lemma 1 is applied.

IV. CASE STUDY

In order to test the ideas presented in this paper simulations of the proposed implementations were performed on a digital computer. The simulation was based on the following differential equations:

$$\frac{dx_k}{dt} = \frac{-1}{\tau}\left(\frac{\partial E_{p,s,c}(x)}{\partial x_k}\right), \quad k = 1,\ldots,n ,$$

where τ is the scaling factor or time constant and $E_{p,s,c}(x)$ is as defined earlier. The benchmark problem which we choose to solve was taken from [2] (see also [4]):

$$\text{minimize } ||x||_p$$
$$\text{subject to } Ax = b ,$$

where $p = 1,2$, or ∞, and

$$A = \begin{bmatrix} 2 & -1 & 4 & 0 & 3 & 1 \\ 5 & 1 & -3 & 1 & 2 & 0 \\ 1 & -2 & 1 & -5 & -1 & 4 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} .$$

The function $E_{p,s,c}(x)$ was constructed for all possible combinations of $s = 1,2$, and ∞ and $p = 1,2$, and ∞. The value of c was computed according to the method given in Proposition 1, that is

$$c = \frac{K}{\sqrt{\lambda_{min}(AA^T)}} .$$

For the problem of interest the value of $\lambda_{min}(AA^T)$ was found to be 24.945. The resulting values of c for each of the nine cases was computed and is shown in the table below

		s		
		1	2	∞
p	1	.490	.490	.849
	2	.200	.200	.347
	∞	.200	.200	.347

Table 1. Values of the constant c for different values of p and s.

For each case we assumed an initial condition $x(0) = 0$. The results of the simulations for $p = 1,2$ and ∞ norms are shown in Figs. 1-9.

For $p = 1$, regardless of the value of s, the trajectory converged to the point (see Fig. 1-3)

$$x = [0.000,\ 0.000,\ 0.192,\ 0.756,\ 0.410,\ 0.000]^T ,$$

which gives $||x||_1 = 1.36$.

For $p = 2$, regardless of the value of s, the trajectory converged to the point (see Fig. 4-6)

$$x = [0.088,\ 0.108,\ 0.273,\ 0.505,\ 0.383,\ -0.310]^T ,$$

which gives $||x||_2 = 0.769$.

For $p = \infty$, regardless of the value of s, the trajectory converged to the point (see Fig. 7-9)

$$x = [0.113,\ 0.372,\ 0.351,\ 0.372,\ 0.372,\ -0.372]^T ,$$

which gives $||x||_\infty = 0.372$.

Thus, for each case the result of the simulations agreed with the analytical solutions to the problem (see eg. [2]).

From the following plots of the trajectories of the variables for the nine cases we can see that the trajectories converged to the solution points within a few time constants.

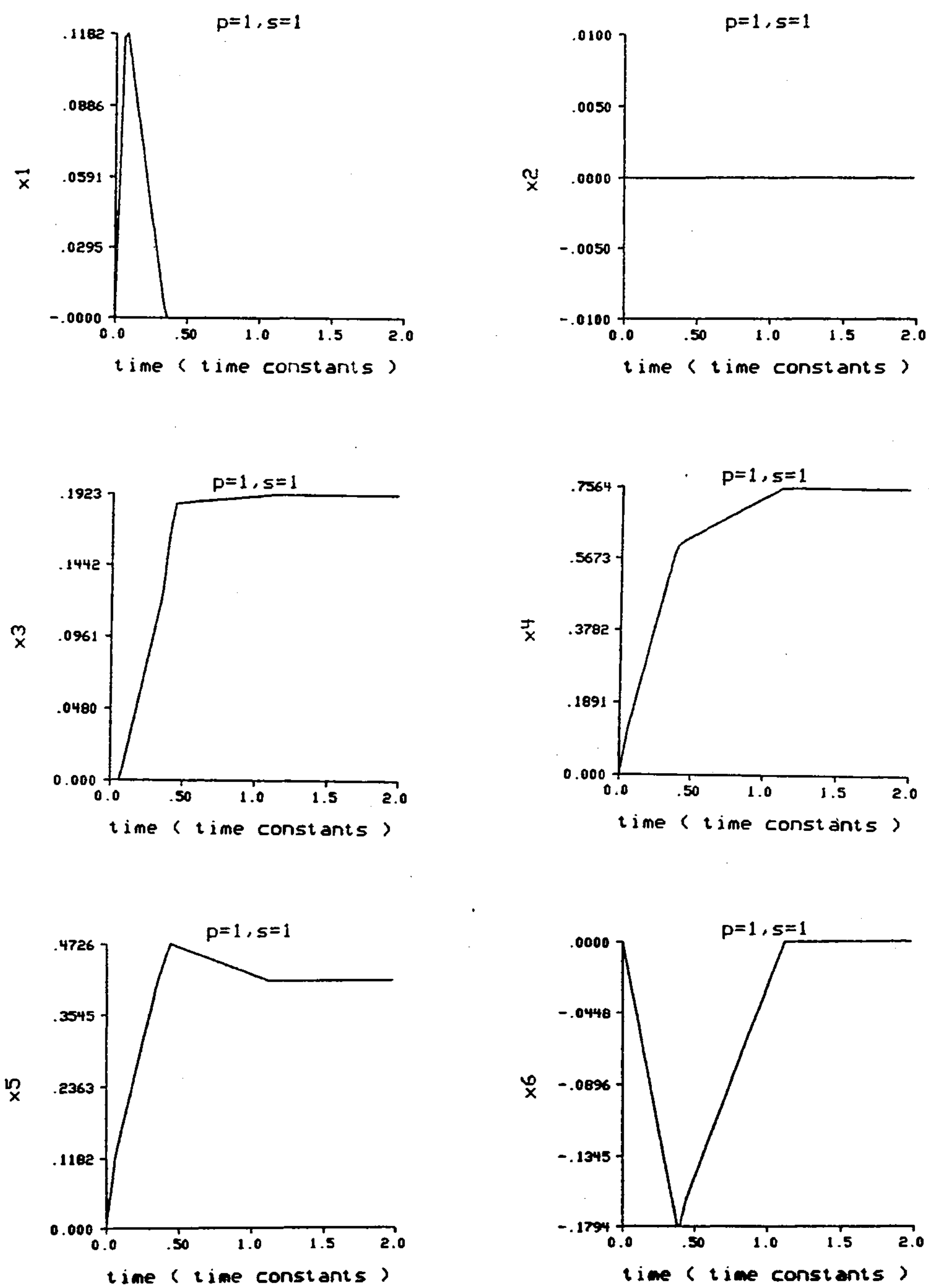

Figure 1. Trajectories corresponding to the case p = 1, s = 1.

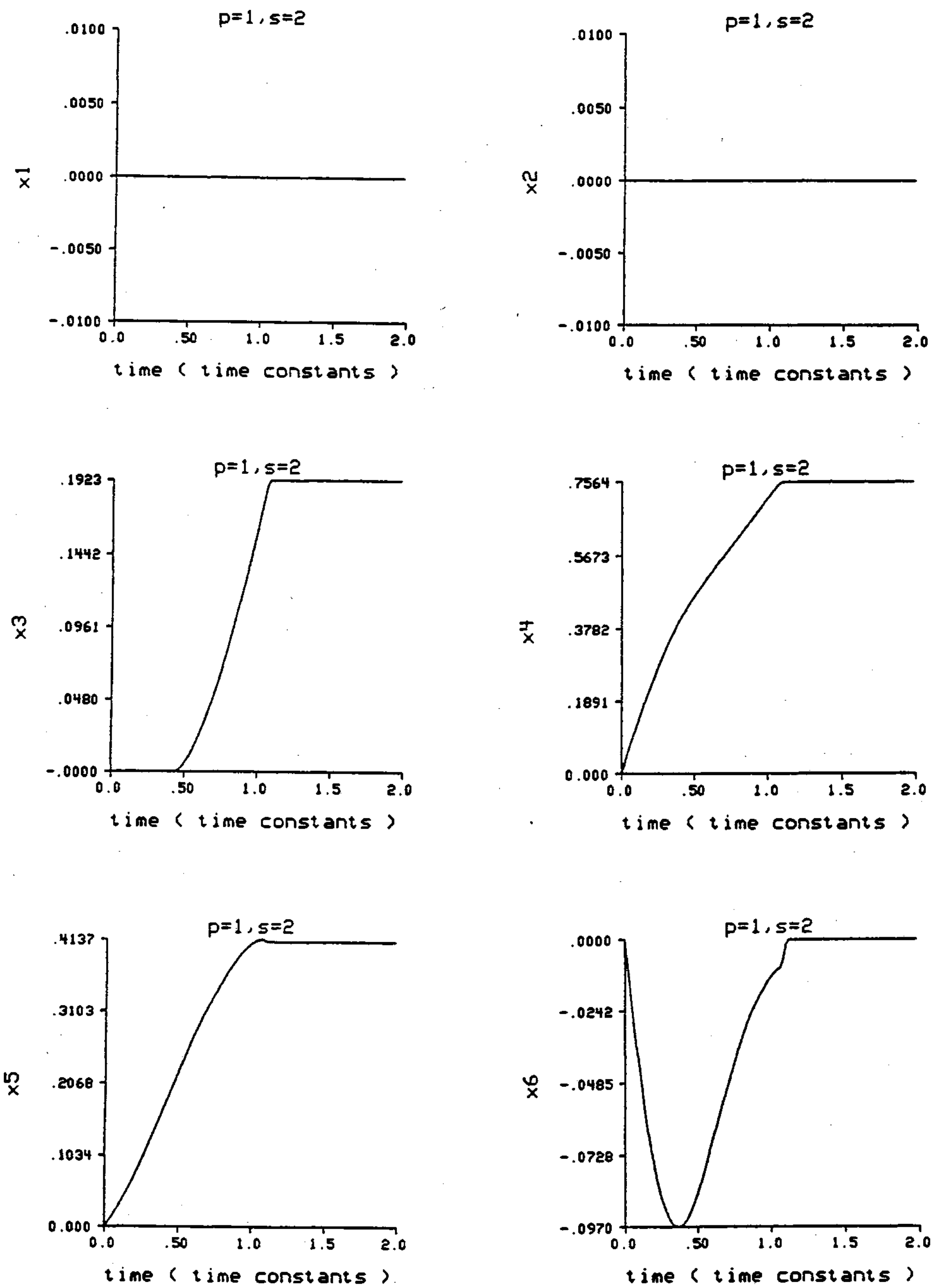

Figure 2. Trajectories corresponding to the case $p = 1$, $s = 2$.

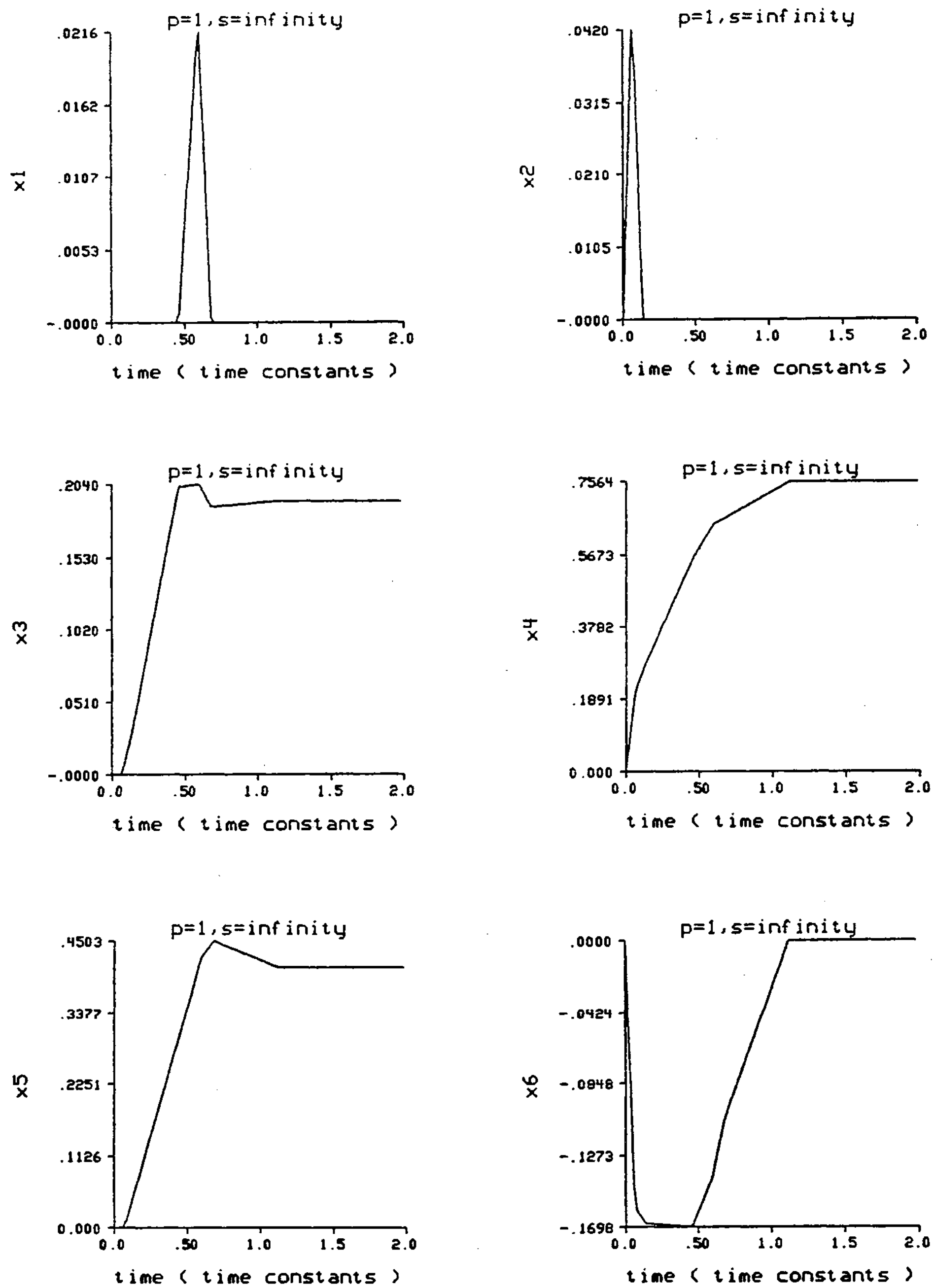

Figure 3. Trajectories corresponding to the case p = 1, s = ∞.

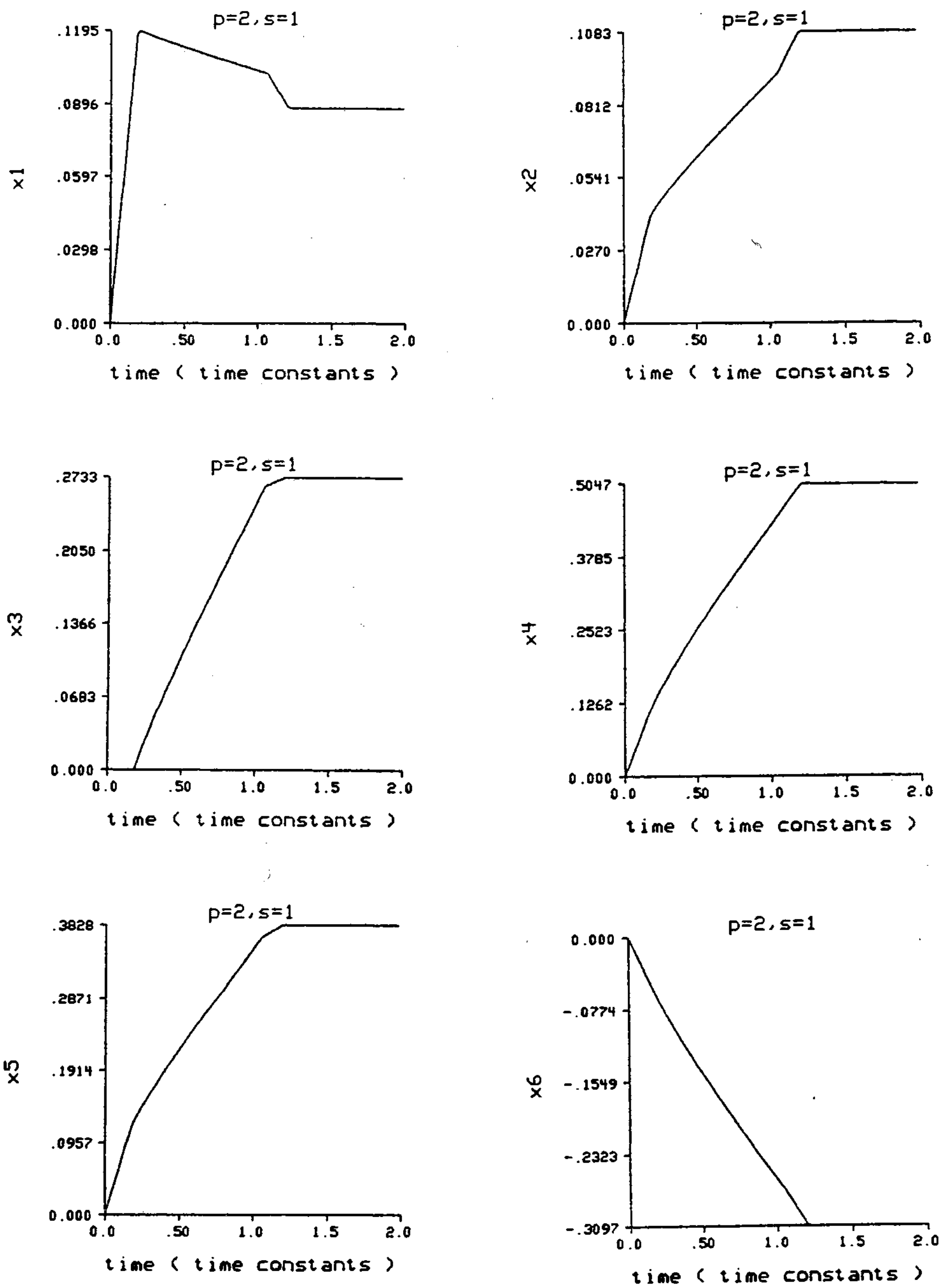

Figure 4. Trajectories corresponding to the case p = 2, s = 1.

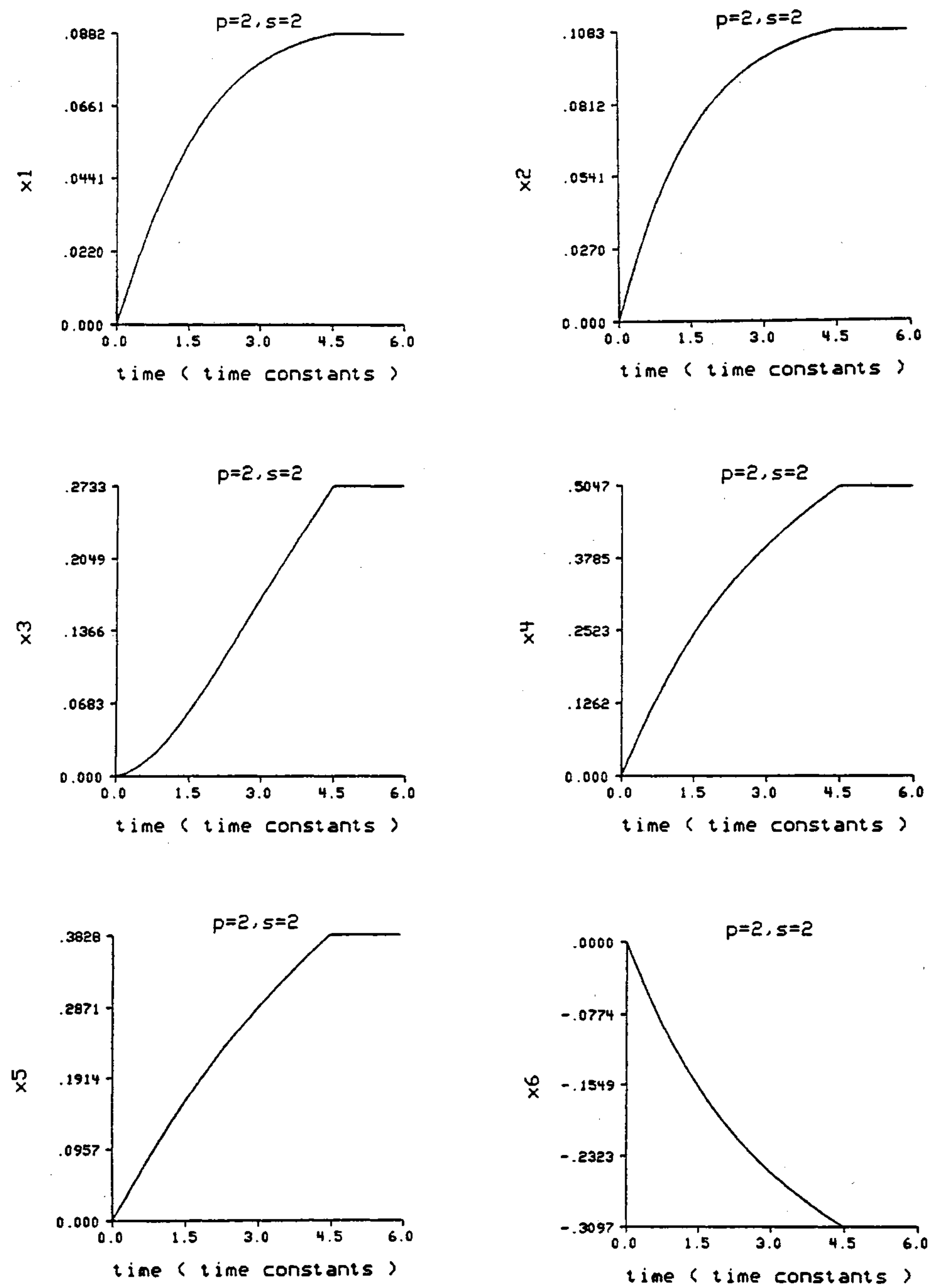

Figure 5. Trajectories corresponding to the case p = 2, s = 2.

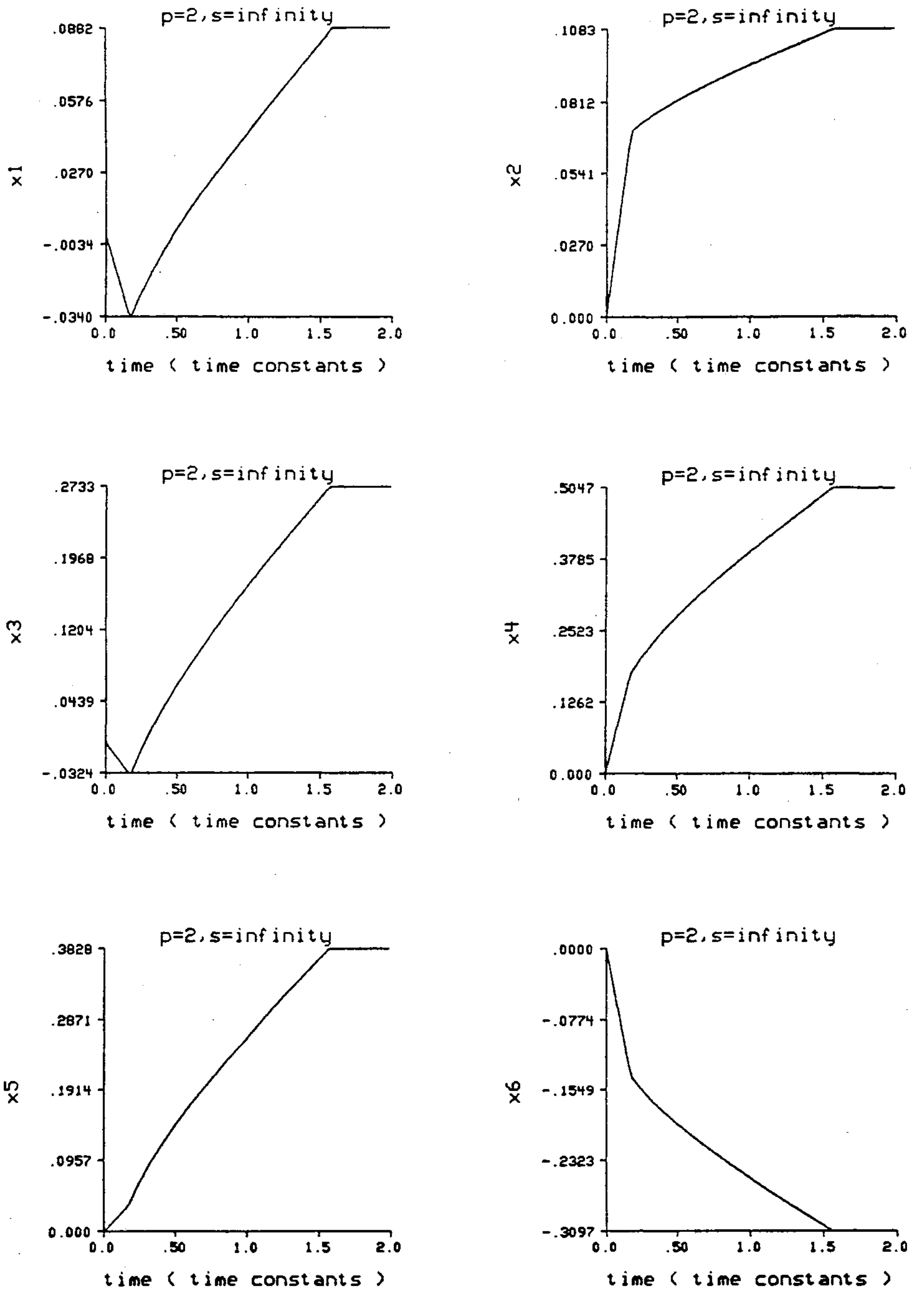

Figure 6. Trajectories corresponding to the case p = 2, s = ∞.

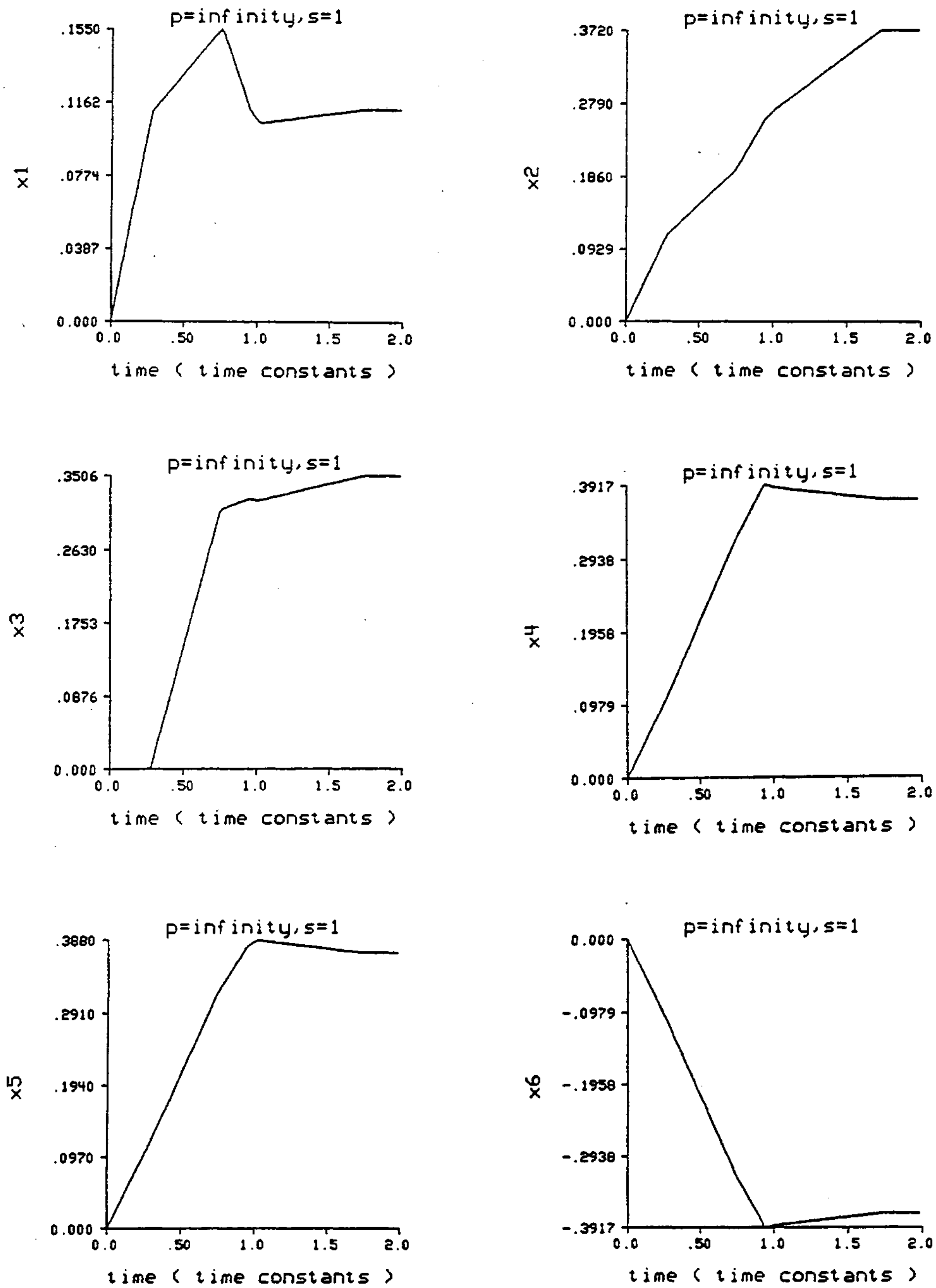

Figure 7. Trajectories corresponding to the case $p = \infty$, $s = 1$.

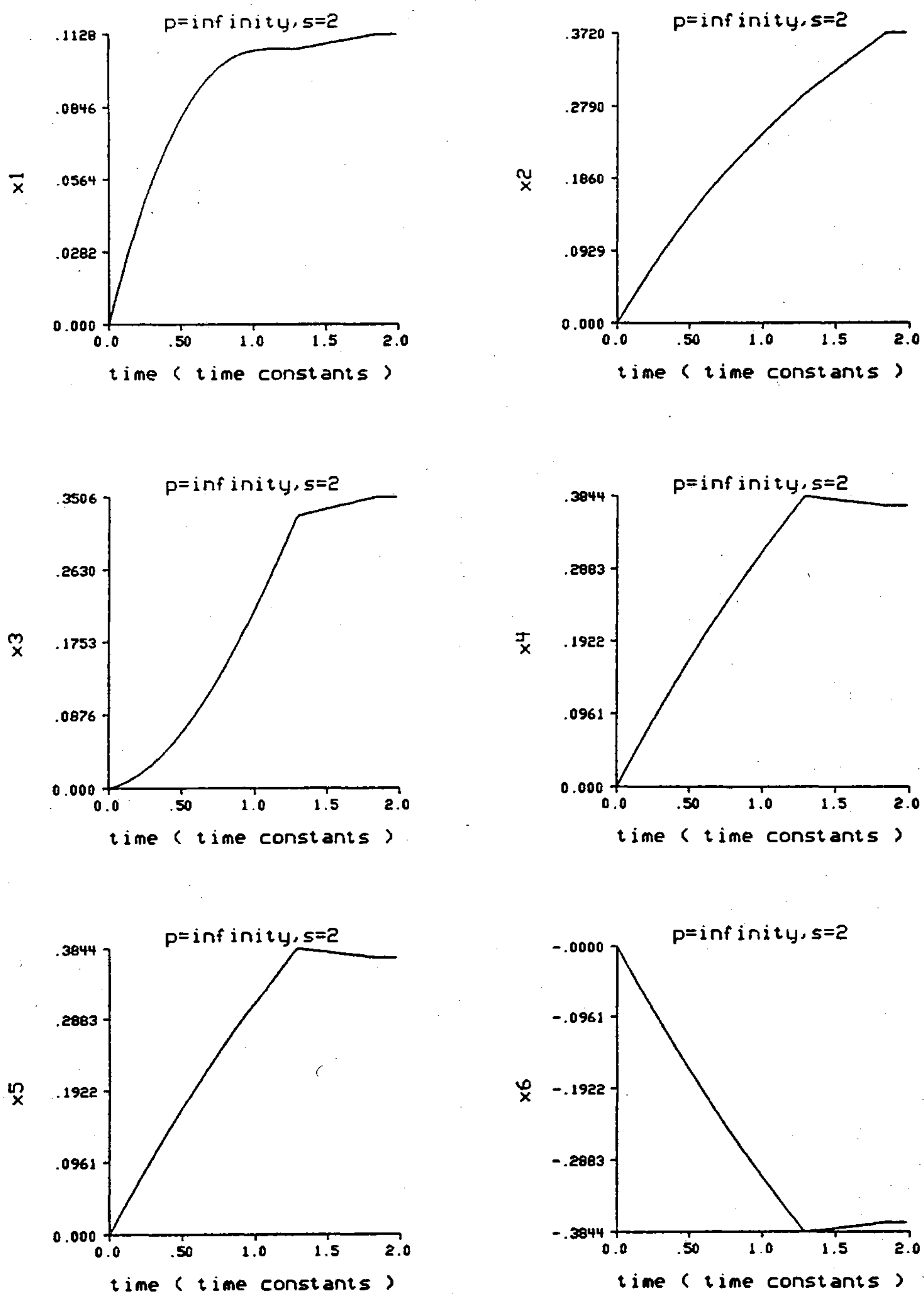

Figure 8. Trajectories corresponding to the case $p = \infty$, $s = 2$.

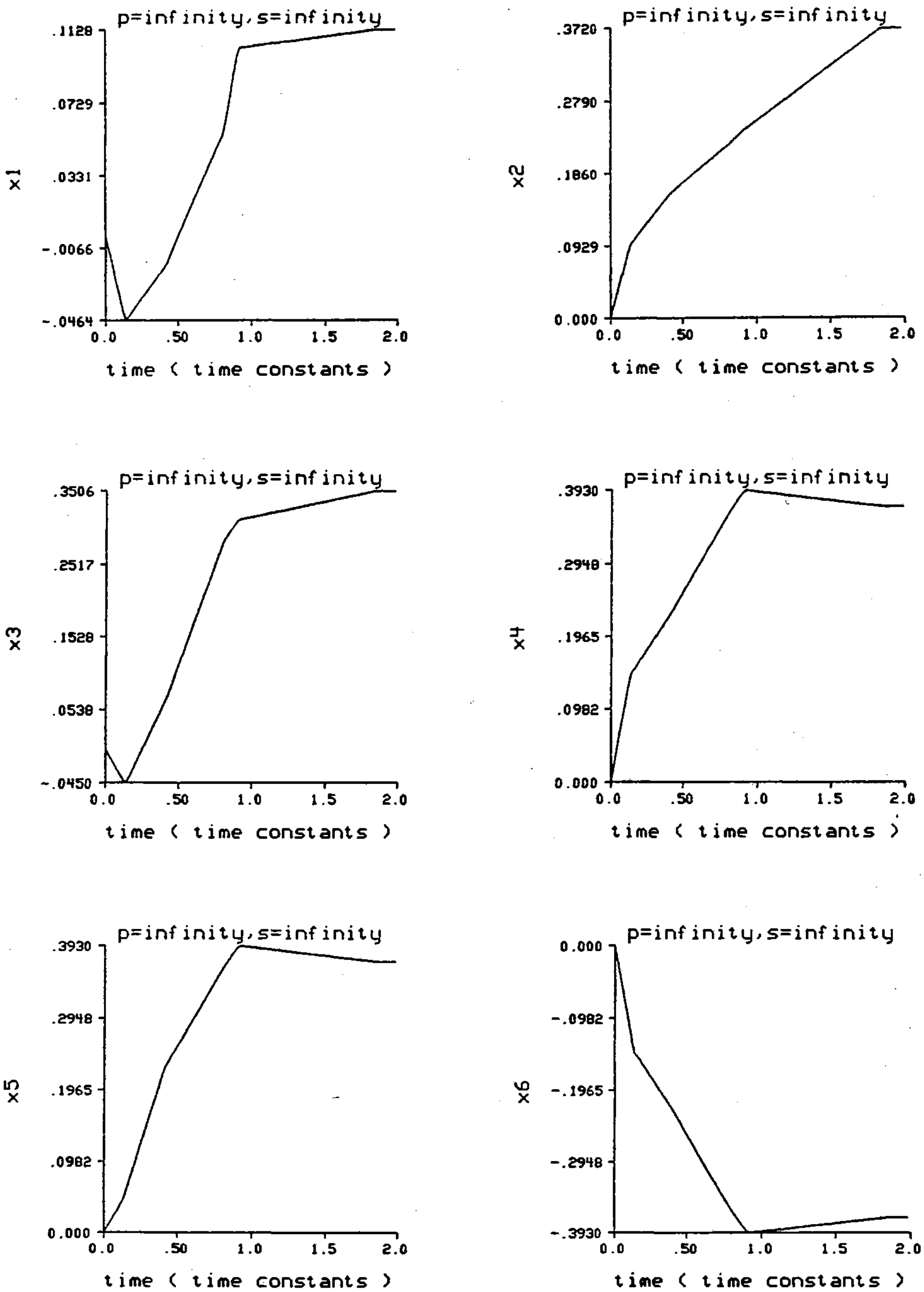

Figure 9. Trajectories corresponding to the case $p = \infty$, $s = \infty$.

V. CONCLUSIONS

A new class of penalty functions has been proposed which enables one to transform constrained optimization problems of quotient space norms minimization to equivalent unconstrained minimization problems. A sharp bound on the weight parameter of the penalty functions has been derived. It has been shown that if the weight parameter is smaller than the derived bound than the constrained and unconstrained problems may not be equivalent. A computationally efficient estimate of the parameter bound has also been given.

REFERENCES

[1] J. A. Cadzow, "Algorithm for the minimum-effort problem." *IEEE Trans. Automatic Control,* Vol. AC-16, No. 1, pp. 60-63, Feb. 1971.

[2] J. A. Cadzow, "Functional analysis and the optimal control of linear discrete systems." *Int. J. Control,* Vol. 17, No. 3, pp. 481-495, 1973.

[3] T. Kailath, "*Linear Systems*", Prentice Hall, Englewood Cliffs, N.J., 1980.

[4] L. Kolev, "Iterative algorithm for the minimum fuel and minimum amplitude problems for linear discrete systems." *Int. J. Control,* Vol. 21, No. 5, pp. 779-784, 1975.

[5] J. P. LaSalle, "*The Stability and Control of Discrete Processes,*" Springer-Verlag, New York, 1986.

[6] D. G. Luenberger, "*Optimization by Vector Space Methods,*" J. Wiley & Sons, New York, 1969.

[7] D. G. Luenberger, "*Linear and Nonlinear Programming,*" (Second Edition), Addison-Wesley Publishing Co., Reading, Mass., 1984.

[8] I. J. Maddox, "*Elements of Functional Analysis,*" (Second Edition), Cambridge Univ. Press, Cambridge, G.B., 1988.