Available online at www.sciencedirect.com
SCIENCE DIRECT
An International Journal
computers & mathematics with applications
ELSEVIER Computers and Mathematics with Applications 48 (2004) 419–427
www.elsevier.com/locate/camwa

Dynamic Programming and Minimal Norm Solutions of Least Squares Problems

R. KALABA AND H. NATSUYAMA
Department of Biomedical Engineering
University of Southern California
Los Angeles, CA 90089, U.S.A.

(Received September 2003; accepted January 2004)

Keywords—Dynamic programming, Least squares, Pseudoinverses.

INTRODUCTION

Least squares problems occur widely in regression analysis, parameter estimation, analytical mechanics, and in many other areas. In [1], we introduced a new dynamic programming approach to least squares problems. The algorithm of that paper relied heavily on knowing the rank of the given matrix and knowing columns which are linearly independent. This paper extends the previous one by removing these restrictions. We develop a new algorithm which we call the $\alpha Q\beta R$ algorithm.

This formulation introduces *two* cost functions, which is new to dynamic programming literature. The first cost function is the square of the length of the current discrepancy vector, and the second is the square of the length of the current solution vector. The two cost functions are to be minimized simultaneously by optimally selecting the minimum length vector solution.

Finally, a connection with Greville's formula for generalized inverses is indicated.

PRINCIPLE OF OPTIMALITY

Let A be an $m \times n$ matrix, b be a column vector of dimension m, and x be a vector of dimension n. Given the matrix A and the vector b, we wish to determine the vector x such that $|Ax - b|^2$ is a minimum and the length of x is as small as possible. There are many approaches to this optimization problem [2]. Here we shall provide an approach through dynamic programming [3]. The reader may wish to consult [4–8].

We introduce *two* cost functions. First we write

$$f_k(b) = \text{the smallest square of the length of the residual vector } A_k x^k - b. \tag{1}$$

Here A_k is a matrix consisting of the first k columns of A and x^k is a column vector of dimension k. We also introduce

$$g_k(b) = \text{the smallest square of the length of the vector } x^k,$$

 Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
doi:10.1016/j.camwa.2004.01.007

where x^k is subject to the restriction

$$\left|A_k x^k - b\right|^2 = \min.\text{over } x^k. \tag{2}$$

In these definitions, $k = 1, 2, 3, \ldots, n$. We now obtain simultaneous recurrence relations for these functions. Having to use two cost functions is curious; yet, it seems unavoidable. We are led to new dynamic programming equations.

Suppose that $f_{k-1}(b)$ and $g_{k-1}(b)$ are known. We wish to obtain $f_k(b)$ and $g_k(b)$. We denote the individual columns of A by $a_1, a_2, \ldots, a_n$. There are two cases to consider, depending on whether a_k is linearly dependent on $a_1, a_2, \ldots, a_{k-1}$ or not. Assume first that a_k is linearly dependent on $a_1, a_2, \ldots, a_{k-1}$. In this case,

$$f_k(b) = f_{k-1}(b), \tag{3}$$

because a linear combination of $a_1, a_2, \ldots, a_{k-1}, a_k$ cannot be brought closer to b than a linear combination of $a_1, a_2, \ldots, a_{k-1}$. We also see that

$$g_k(b) =_{x_k}^{\min} \left[x_k^2 + g_{k-1}(b - x_k a_k)\right], \tag{4}$$

where x_k is a scalar. This follows because if x_k is the k^{th} component of x^k, then the term in square brackets is the square of x_k plus the smallest square of the length of a vector x^{k-1}, where $|A_{k-1}x^{k-1} - (b - x_k a_k)|^2 = \min.\text{over } x^{k-1}$.

Next we assume that a_k is linearly independent of the vectors $a_1, a_2, \ldots, a_{k-1}$. In this case, we must choose x_k, the k^{th} component of x^k in the approximation of b by $A_k x^k$, so that we minimize $f_{k-1}(b - x_k a_k)$. The reason is that with any choice of the scalar x_k, we must approximate the new target vector $b - x_k a_k$ as well as possible through choice of the sum $x_1 a_1 + \cdots + x_{k-1} a_{k-1}$. Thus, we may write

$$f_k(b) =_{x_k}^{\min} f_{k-1}(b - x_k a_k). \tag{5}$$

If the minimizing value of the scalar x_k is x_k^*, then we also have

$$g_k(b) = \left(x_k^*\right)^2 + g_{k-1}\left(b - x_k^* a_k\right). \tag{6}$$

Equations (3)–(6) constitute the desired system of recurrence relations. Equations (3) and (4) apply if a_k is dependent on $a_1, a_2, \ldots, a_{k-1}$, and equations (5) and (6) apply if a_k is independent of the earlier columns of the matrix A. The underlying role of Bellman's principle of optimality is clear [3].

In addition, for the case $k = 1$, we have

$$f_1(b) =_{x_1}^{\min} (a_1 x_1 - b)^\top (a_1 x_1 - b), \qquad a \neq 0. \tag{7}$$

Thus,

$$f_1(b) =_{x_1}^{\min} \left[a_1^\top a_1 x_1^2 - 2a_1^\top b x_1 + b^\top b\right]. \tag{8}$$

The minimizing condition is

$$a_1^\top a_1 x_1 - a_1^\top b = 0, \tag{9}$$

so that

$$x_1^* = \frac{a_1^\top b}{a_1^\top a_1} = a_1^+ b. \tag{10}$$

Here we have used the fact that the generalized inverse of the vector a_1, a_1^+, is $a_1^\top / a_1^\top a_1$ (assuming that $a_1 \neq 0$). It follows that

$$\begin{aligned} f_1(b) &= \left[b^\top \left(a_1^+\right)^\top a_1^\top a_1 a_1^+ b - 2b^\top a_1 a_1^+ b + b^\top b\right], \\ f_1(b) &= b^\top \left[I - a_1 a_1^+\right] b. \end{aligned} \tag{11}$$