COMMENTS ON “JORDAN CANONICAL FORM
OF THE GOOGLE MATRIX”∗

GANG WU† AND YIMIN WEI‡

Abstract. The Google matrix is a Web hyperlink matrix which is given by $P(\alpha) = \alpha P + (1-\alpha)E$, where $P$ is a row stochastic matrix, $E$ is a row stochastic rank-one matrix, and $0 < \alpha < 1$. In this paper we explore the analytic expression of the Jordan canonical form and point out that a theorem due to Serra-Capizzano (cf. Theorem 2.3 in [SIAM J. Matrix Anal. Appl., 27 (2005), pp. 305–312]) can be used for estimating the condition number of the PageRank vector as a function of $\alpha$ now viewed in the complex field. Furthermore, we give insight into a more efficient scaling matrix in order to minimize the condition number.

Key words. PageRank, Google matrix, Jordan canonical form, condition number

AMS subject classifications. 65F15, 65F10, 65C40

DOI. 10.1137/070682204

1. Preliminaries. An important problem in Web searches is determining the importance of each page. The major ingredient in determining the order to display Web pages is PageRank [4]. The PageRank vector is the stationary distribution of the Google matrix, a stochastic and irreducible matrix whose dimension can reach $10^9$ [1, 7, 11].

Analysis of the PageRank formula provides an interesting topic for the PageRank problem [6, 8, 9, 10]. Recently, Horn and Serra-Capizzano [6] and Serra-Capizzano [8, 9] determined the analytic expression of the Jordan canonical form of the Google matrix. Theorem 2.3 in [8] (see also Theorem 8.2 in [9]) is quoted as follows, which depicts the eigenvalues and invariant subspace of the Google matrix.

**Theorem 1.** Let $P$ be a row stochastic matrix of size $n$, let $\alpha \in (0, 1)$, and let $E = ev^H$ be a row stochastic rank one matrix of size $n$ with $e$ the vector of all ones and with $v$ an $n$-sized vector representing a probability distribution, i.e., $v_i \geq 0$ and $||v||_1 = 1$. Consider the matrix $P(\alpha) = \alpha P + (1-\alpha)E$ and let $P = XJ(\alpha)X^{-1}$, $X = [e, x_2, \ldots, x_n]$, $Y = (X^{-1})^H = [y_1, y_2, \ldots, y_n]$,

$$J(\alpha) = \begin{bmatrix}
1 & \alpha \cdot & * \\
\alpha \lambda_2 & \alpha \cdot & * \\
\vdots & \ddots & \ddots \\
\alpha \lambda_{n-1} & \alpha \cdot & * \\
\alpha \lambda_n & & \\
\end{bmatrix},$$

Received by the editors February 8, 2007; accepted for publication (in revised form) by D. P. O’Leary September 4, 2007; published electronically April 23, 2008.

http://www.siam.org/journals/simax/30-1/68220.html

†School of Mathematical Sciences, Xuzhou Normal University, Xuzhou, 221116, Jiangsu, People’s Republic of China (gangwu76@yahoo.com.cn, wugangzy@gmail.com). This author was supported by the National Science Foundation of China under grant 10626044 and by the Qing-Lan Project of Jiangsu Province.

‡Corresponding author. School of Mathematical Sciences and Key Laboratory of Mathematics for Nonlinear Sciences, Fudan University, Shanghai, 200433, People’s Republic of China (ymwei@fudan.edu.cn, ymwei.cn@yahoo.com). This author was supported by the National Science Foundation of China under grant 10471027 and by the Shanghai Education Committee.
and

\[
J(\alpha) = D^{-1} = \begin{bmatrix}
1 & \alpha \lambda_2 & * \\
& \ddots & \ddots \\
& & \alpha \lambda_{n-1} & * \\
& & & \alpha \lambda_n
\end{bmatrix} \quad D, \quad D = \text{diag}(1, \alpha, \ldots, \alpha^{n-1}),
\]

with * denoting a value that can be 0 or 1. Then we have

\[
P(\alpha) = ZJ(\alpha)Z^{-1}, \quad Z = XR^{-1},
\]

and, in addition, the following facts hold.

- \(1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|\) and \(\lambda_2 = 1\) if \(P\) is reducible and its graph has at least two irreducible closed sets.
- We have

\[
R = I + e_1 w^H, \quad w^H = (0, w_2, \ldots, w_n),
\]

where

\[
w_2 = (1 - \alpha)v^H x_2/(1 - \alpha \lambda_2),
\]

\[
w_j = ((1 - \alpha)v^H x_j + [J(\alpha)]_{j-1,j} w_{j-1})/(1 - \alpha \lambda_j), \quad j = 3, 4, \ldots, n.
\]

We mention that in the original paper by Serra-Capizzano, there is a typo since \(D\) and \(D^{-1}\) are exchanged in (1). The following corollary due to Serra-Capizzano [8, Corollary 2.4] gives an analytic expression of the PageRank vector.

**Corollary 2.** With the notation of Theorem 1, the PageRank vector \([y^I(\alpha)]^H\) is given by

\[
[y^I(\alpha)]^H = y_1^H + \sum_{j=2}^n w_j y_j^H.
\]

As was pointed out in [8], a strong challenge posed by formula (3) is the possibility of using vector extrapolation for obtaining the expression of \([y^I(1)]^H\). The idea of the extrapolation procedure is to start from values of \([y^I(\alpha)]^H\) for some different values of \(\alpha\) (possibly far from 1), then to compute the unknowns appearing in (3), and finally to compute \([y^I(1)]^H\). This subject is under investigation in [2, 3].

**2. On the condition number of the PageRank vector.** Recall from (1) and (2) that the PageRank vector \([y^I(\alpha)]^H\) is the first row of the matrix \(DZ^{-1} = DRX^{-1}\), that is, \([y^I(\alpha)]^H = e_1^H(DRX^{-1})\). If we denote \(W = XR^{-1}D^{-1}\), then formula (3) can be rewritten as

\[
[y^I(\alpha)]^H \cdot W = e_1^H.
\]

It is well known that the sensitivity of the linear system (4) is closely related to the condition number \(\kappa(W)\) of \(W\) [5], where

\[
\kappa(W) = \frac{||W|| \cdot ||W^{-1}||}{\text{min}_i \frac{1}{\lambda_i(W)}} = \frac{||XR^{-1}D^{-1}|| \cdot ||DX^{-1}||}{\kappa(X) \cdot \kappa(DR)}.
\]
Therefore, $W$ will be ill-conditioned, provided either $X$ or $DR$ is ill-conditioned. We have the following theorem on the conditioning of $DR$ with respect to $\infty$-norm.

**Theorem 3.** Under the above notation, if $D = \text{diag}(1, \alpha, \ldots, \alpha^{n-1})$ and $0 < \alpha < 1$, then

\[
\kappa_\infty(DR) = \max \left\{ \left(1 + \sum_{j=2}^{n} |w_j| \right) \left(1 + \sum_{j=2}^{n} \frac{|w_j|}{\alpha^{j-1}}\right), \frac{1}{\alpha^{n-1}} \left(1 + \sum_{j=2}^{n} |w_j| \right) \right\},
\]

where

\[
|w_2| = \frac{(1-\alpha)|v^H x_2|}{|1-\alpha \lambda_2|},
\]

and

\[
|w_j| \leq \alpha^{j-2} \cdot \frac{(1-\alpha)|v^H x_2|}{|(1-\alpha \lambda_2) \cdots (1-\alpha \lambda_j)|} + \alpha^{j-3} \cdot \frac{(1-\alpha)|v^H x_3|}{|(1-\alpha \lambda_3) \cdots (1-\alpha \lambda_j)|} + \cdots + \frac{(1-\alpha)|v^H x_j|}{|1-\alpha \lambda_j|}, \quad j = 3, 4, \ldots, n.
\]

Specifically, if $P$ is diagonalizable, then

\[
\kappa_\infty(DR) \geq \left(1 + \sum_{j=2}^{n} \frac{(1-\alpha)|v^H x_j|}{|1-\alpha \lambda_j|} \right) \left(1 + \sum_{j=2}^{n} \frac{(1-\alpha) \cdot |v^H x_j|}{\alpha^{j-1} \cdot |1-\alpha \lambda_j|} \right).
\]

**Proof.** Since

\[
R = \begin{bmatrix} 1 & w_2 & \cdots & w_n \\ \vdots & 1 & \ddots & \vdots \\ & & \ddots & 1 \\ 1 & & & 1 \end{bmatrix},
\]

it is easy to verify that

\[
R^{-1} = \begin{bmatrix} 1 & -w_2 & \cdots & -w_n \\ \vdots & 1 & \ddots & \vdots \\ & & \ddots & 1 \\ 1 & & & 1 \end{bmatrix}, \quad DR = \begin{bmatrix} 1 & w_2 & \cdots & w_n \\ \alpha & \alpha^2 & \cdots & \alpha^{n-1} \end{bmatrix},
\]

and

\[
R^{-1} D^{-1} = \begin{bmatrix} 1 & -w_2/\alpha & -w_3/\alpha^2 & \cdots & -w_n/\alpha^{n-1} \\ 1/\alpha & 1/\alpha^2 & \cdots & 1/\alpha^{n-1} \end{bmatrix}.
\]

Therefore,

\[
||DR||_\infty = 1 + \sum_{j=2}^{n} |w_j|, \quad ||R^{-1} D^{-1}||_\infty = \max \left\{ 1 + \sum_{j=2}^{n} \frac{|w_j|}{\alpha^{j-1}}, \frac{1}{\alpha^{n-1}} \right\}.
\]
and
\[
\kappa_\infty(DR) = ||DR||_\infty ||R^{-1}D^{-1}||_\infty \\
= \max \left\{ \left( 1 + \sum_{j=2}^{n} |w_j| \right) \left( 1 + \sum_{j=2}^{n} \frac{|w_j|}{\alpha^{j-1}} \right), \frac{1}{\alpha^{n-1}} \left( 1 + \sum_{j=2}^{n} |w_j| \right) \right\}.
\]

Recall from Theorem 1 that
\[
w_2 = \frac{(1 - \alpha)v^H_2}{1 - \alpha \lambda_2}, \quad w_3 = \frac{(1 - \alpha)v^H_3 + (\alpha \cdot \ast)w_2}{1 - \alpha \lambda_3},
\]
where \( \ast \) stands for either 0 or 1. Notice that
\[
|w_3| \leq \frac{(1 - \alpha)|v^H_3| + \alpha \cdot |w_2|}{|1 - \alpha \lambda_3|} = \alpha \cdot \frac{(1 - \alpha)|v^H_2|}{|1 - \alpha \lambda_2|} \cdot \frac{|v^H_3|}{|1 - \alpha \lambda_3|} + (1 - \alpha) \cdot \frac{|v^H_3|}{|1 - \alpha \lambda_3|}.
\]

Inductively, suppose that
\[
|w_{j-1}| \leq \alpha^{j-3} \cdot \frac{(1 - \alpha)|v^H_2|}{|1 - \alpha \lambda_2| \cdots (1 - \alpha \lambda_{j-1})} + \alpha^{j-4} \cdot \frac{(1 - \alpha)|v^H_3|}{|1 - \alpha \lambda_3| \cdots (1 - \alpha \lambda_{j-1})} \cdots + \alpha^{j-4} \cdot \frac{(1 - \alpha)|v^H_{j-1}|}{|1 - \alpha \lambda_{j-1}|}.
\]

From Theorem 1, we obtain
\[
w_j = \frac{((1 - \alpha)v^H_j + (\alpha \cdot \ast)w_{j-1})}{1 - \alpha \lambda_j},
\]
so
\[
|w_j| \leq \frac{(1 - \alpha)|v^H_j| + \alpha \cdot |w_{j-1}|}{|1 - \alpha \lambda_j|} \leq \alpha^{j-2} \cdot \frac{(1 - \alpha)|v^H_2|}{|1 - \alpha \lambda_2| \cdots (1 - \alpha \lambda_j)} + \alpha^{j-3} \cdot \frac{(1 - \alpha)|v^H_3|}{|1 - \alpha \lambda_3| \cdots (1 - \alpha \lambda_j)} \cdots + \alpha^{j-4} \cdot \frac{(1 - \alpha)|v^H_{j-1}|}{|1 - \alpha \lambda_{j-1}|}.
\]

Specifically, when \( P \) is diagonalizable, it follows from (5) that
\[
\kappa_\infty(DR) \geq \left( 1 + \sum_{j=2}^{n} |w_j| \right) \left( 1 + \sum_{j=2}^{n} \frac{|w_j|}{\alpha^{j-1}} \right),
\]
and recall from Theorem 2.1 in [8] that
\[
|w_j| = \frac{(1 - \alpha)|v^H_j|}{|1 - \alpha \lambda_j|}, \quad j = 2, 3, \ldots, n,
\]
and (6) is obtained from combining (7) and (8). \( \square \)

Theorem 3 indicates that \( DR \) may be ill-conditioned as the number \( n \) in (5) and (6) is often very huge, being the total number of Web pages (in millions or billions).
Consequently, $W$ can be ill-conditioned in practice. Actually, one is not recommended to use (3) directly. One reason is that the dimension $n$ is large, and the expression in (3) is simplified by replacing $n$ with a much smaller value $m$ [3]. Theorem 3 gives another reason: as we have just observed, $W$ may be ill-conditioned even if $X$ is well-conditioned and $\alpha$ is far from 1, which implies that a small change in $W$ can give a dramatic change in the PageRank vector.

However, we would like to point out that the results presented in Theorem 3 are not so strong. In Theorem 1, the matrix $D$ is chosen as diag$(1, \alpha, \ldots, \alpha^{n-1})$. In fact, the scaling matrix is not unique. For instance, if we choose $\hat{D} = \text{diag}(1, \alpha^{-1}, \ldots, \alpha^{-1})$, then $D \lambda D^{-1} = \hat{D} \lambda \hat{D}^{-1}$. So it is interesting to take into account the more efficient scaling matrix $D$ (which is not unique at all) in order to decrease the estimate of the condition number discussed in Theorem 3.

3. How to use clever choices of the scaling matrix. The conditioning for nonnegative $\alpha$ less than one is known to be bounded by $2/(1 - \alpha)$ [7]. Therefore the interest of this paper is for $\alpha$ outside the unit cycle (i.e., $\alpha \in C$ and $|\alpha| > 1$ [6, 9]), and to find interesting results one should use clever choices of the scaling matrix $D$ that minimizes the condition number. That is, we consider how to define a new matrix $\hat{D}$ such that

$$\kappa_\infty(\hat{D}R) = \min_{\hat{D} \text{ diagonal and nonsingular}} \kappa_\infty(DR),$$

with the constraint that

$$DJ(\alpha)D^{-1} = \hat{D}J(\alpha)\hat{D}^{-1} = \hat{J},$$

where $\hat{J}$ is the Jordan canonical form of $P(\alpha)$.

However, determining an optimal matrix of minimal conditioning is a very complicated task, and the result is problem dependent. In this paper we give insight into three special cases, and the results extend easily to cover the general case, at the cost of much heavier notation. We assume from now on that $\alpha \in C$, $|\alpha| < 1$, and $1 - \alpha\lambda_j \neq 0$ ($j = 2, 3, \ldots, n$) so that $w_j$ can be well defined; see Theorem 1. Furthermore, we emphasize that all the analysis given below also applies to the case when $\alpha \in C$, $|\alpha| < 1$.

Let $1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $P$. Then, we have from Theorem 1 that the eigenvalues of $P(\alpha)$ are $1, \alpha \lambda_2, \ldots, \alpha \lambda_n$. Suppose that $|\lambda_2| = |\lambda_3| = \cdots = |\lambda_p| = 1$, and $|\lambda_j| < 1$, $j \geq p + 1$. It follows from Theorem 8.2 (ii) of [6] (see also Theorem 7.2 (ii) of [9]) that $\alpha \lambda_2, \alpha \lambda_3, \ldots, \alpha \lambda_p$ are (semi)simple eigenvalues of $P(\alpha)$, thus the Jordan canonical form of $P(\alpha)$ takes the form

$$\hat{J} = \begin{bmatrix} 1 \oplus [\alpha \lambda_2] \oplus \cdots \oplus [\alpha \lambda_p] \oplus J_{n_1}(\alpha \nu_1) \oplus \cdots \oplus J_{n_k}(\alpha \nu_k) \end{bmatrix}$$

and

$$J(\alpha) = \begin{bmatrix} 1 \oplus [\alpha \lambda_2] \oplus \cdots \oplus [\alpha \lambda_p] \oplus \alpha J_{n_1}(\nu_1) \oplus \cdots \oplus \alpha J_{n_k}(\nu_k) \end{bmatrix},$$

where

$$J_{n_i}(\nu_i) = \begin{bmatrix} \nu_i & * & \cdots & \cdots & \nu_i \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \nu_i & * \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \nu_i \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \nu_i \end{bmatrix} \in C^{n_i \times n_i}, \quad i = 1, 2, \ldots, k,$$

and $\{\nu_1, \nu_2, \ldots, \nu_k\} \subset \{\lambda_2, \lambda_3, \ldots, \lambda_n\}$. 
Case 1. $J_n((\alpha_1, \ldots, \alpha_k)) = \text{diag}(\alpha_1, \ldots, \alpha_k)$, $i = 1, 2, \ldots, k$.

In this case, $P(\alpha)$ is diagonalizable, and for any nonsingular diagonal matrix $D$
there holds $J(\alpha) = D^{-1}JD$. We have the following theorem.

**Theorem 4.** If $P(\alpha)$ is diagonalizable, then the optimal scaling matrix can be
chosen as

$$
\tilde{D} = \text{diag}\left(1, 1 + \sum_{j=2}^{n} |w_j|, \ldots, 1 + \sum_{j=2}^{n} |w_j|\right),
$$

and the minimal condition number is

$$
\kappa_\infty(\tilde{D}R) = 1 + 2 \sum_{j=2}^{n} |w_j|.
$$

**Proof.** Without loss of generality, let $D = \text{diag}(1, d_2, \ldots, d_n)$ with $d_j \neq 0$, $j = 2, 3, \ldots, n$. So we have

$$
DR = 
\begin{bmatrix}
1 & w_2 & \cdots & w_n \\
& d_2 & \ddots \\
& & \ddots & d_n \\
& & & 1/d_n
\end{bmatrix},
$$

and

$$
||DR||_\infty = \max\left\{1 + \sum_{j=2}^{n} |w_j|, \max_{2 \leq j \leq n} |d_j|\right\}.
$$

On the other hand,

$$
R^{-1}D^{-1} = 
\begin{bmatrix}
1 & -w_2/d_2 & \cdots & -w_n/d_n \\
& 1/d_2 & \ddots \\
& & \ddots & 1/d_n \\
& & & 1/d_n
\end{bmatrix},
$$

and

$$
||R^{-1}D^{-1}||_\infty = \max\left\{1 + \sum_{j=2}^{n} |w_j/d_j|, \frac{1}{\min_{2 \leq j \leq n} |d_j|}\right\}.
$$

(a) If $\max_{2 \leq j \leq n} |d_j| \leq 1 + \sum_{j=2}^{n} |w_j|$, then

$$
\kappa_\infty(DR) = \max\left\{\left(1 + \sum_{j=2}^{n} |w_j|\right)\left(1 + \sum_{j=2}^{n} |w_j/d_j|\right), \frac{1 + \sum_{j=2}^{n} |w_j|}{\min_{2 \leq j \leq n} |d_j|}\right\}.
$$

So as to minimize the condition number, we have to pick $\min_{2 \leq j \leq n} |d_j|$ as large as
possible. As a result, $\min_{2 \leq j \leq n} |d_j| = \max_{2 \leq j \leq n} |d_j| = 1 + \sum_{j=2}^{n} |w_j|$ is a reasonable
choice.

(b) If $\max_{2 \leq j \leq n} |d_j| \geq 1 + \sum_{j=2}^{n} |w_j|$, then

$$
\kappa_\infty(DR) = \max\left\{\frac{\max_{2 \leq j \leq n} |d_j|}{\min_{2 \leq j \leq n} |d_j|}\left(1 + \sum_{j=2}^{n} |w_j/d_j|\right), \frac{\max_{2 \leq j \leq n} |d_j|}{\min_{2 \leq j \leq n} |d_j|}\right\}.
$$
So as to minimize the condition number, we have to set \( \max_{2 \leq j \leq n} |d_j| \) and \( \frac{\max_{2 \leq j \leq n} |d_j|}{\min_{2 \leq j \leq n} |d_j|} \) as small as possible. As a result, \( \min_{2 \leq j \leq n} |d_j| = \max_{2 \leq j \leq n} |d_j| = 1 + \sum_{j=2}^{n} |w_j| \) is a reasonable choice, and (14) is a direct conclusion from (15) or (16).

Case 2.

\[
J_{n_1} (\alpha \nu_i) = \begin{bmatrix} \alpha \nu_i & 1 \\ \vdots & \ddots & \ddots \\ \alpha \nu_i & & 1 \\ \alpha \nu_i & & & 1 \end{bmatrix} \in \mathbb{C}^{n_1 \times n_1}, \quad i = 1, 2, \ldots, k.
\]

In this case, all the eigenvalues \( \alpha \nu_i \) \((i = 1, 2, \ldots, k)\) of \( P(\alpha) \) are defective, where \( \{\nu_1, \nu_2, \ldots, \nu_k\} \subset \{\lambda_{p+1}, \lambda_{p+2}, \ldots, \lambda_{n}\} \), and

\[
J(\alpha) = \text{diag}(1, \alpha \lambda_2, \ldots, \alpha \lambda_p) \oplus \begin{bmatrix} \alpha \lambda_{p+1} & \alpha \\ \vdots & \ddots \\ \alpha \lambda_{n-1} & \alpha \\ \alpha \lambda_n & \end{bmatrix}
\]

and

\[
\hat{J} = \text{diag}(1, \alpha \lambda_2, \ldots, \alpha \lambda_p) \oplus \begin{bmatrix} \alpha \lambda_{p+1} & 1 \\ \vdots & \ddots \\ \alpha \lambda_{n-1} & 1 \\ \alpha \lambda_n & \end{bmatrix}.
\]

We consider the scaling matrix of the form \( D = \text{diag}(1, \ldots, 1, \delta_{p+1}, \ldots, \delta_n) \in \mathbb{C}^{n \times n} \). The following lemma gives a necessary and sufficient condition for the matrix \( D \) satisfying \( J(\alpha) = D^{-1} \hat{J} D \).

**Lemma 5.** Let \( D = \text{diag}(1, \ldots, 1, \delta_{p+1}, \ldots, \delta_n) \in \mathbb{C}^{n \times n} \) be any nonsingular matrix. Then

\[
J(\alpha) = D^{-1} \hat{J} D \iff \delta_j = \alpha \delta_{j-1}, \quad j = p + 2, p + 3, \ldots, n.
\]

**Proof.** Note that \( J(\alpha) = D^{-1} \hat{J} D \iff DJ(\alpha) = \hat{J} D \). On the one hand,

\[
D \cdot J(\alpha) = \text{diag}(1, \alpha \lambda_2, \ldots, \alpha \lambda_p) \oplus \begin{bmatrix} \alpha \delta_{p+1} & \alpha \delta_{p+1} \\ \alpha \delta_{p+2} & \alpha \delta_{p+2} \\ \vdots & \ddots \\ \alpha \delta_{n-1} & \alpha \delta_{n-1} \\ \alpha \delta_n & \alpha \delta_n \end{bmatrix}.
\]

On the other hand,

\[
\hat{J} \cdot D = \text{diag}(1, \alpha \lambda_2, \ldots, \alpha \lambda_p) \oplus \begin{bmatrix} \alpha \delta_{p+1} & \delta_{p+2} \\ \alpha \delta_{p+1} & \delta_{p+2} \\ \vdots & \ddots \\ \alpha \delta_{n-1} & \delta_{n-1} \\ \alpha \delta_n & \delta_n \end{bmatrix},
\]

and (17) is obtained trivially by comparing the superdiagonal of the two matrices. \( \square \)
Therefore, it follows from Lemma 5 that $D$ takes the form
\begin{equation}
D = \text{diag}(1, \ldots, 1, \delta_{p+1}, \alpha \delta_{p+1}, \ldots, \alpha^{n-p-1} \delta_{p+1}), \quad \delta_{p+1} \neq 0,
\end{equation}
and the problem of defining the optimal matrix $\tilde{D}$ resorts to determining an appropriate value $\delta_{p+1}$.

**Theorem 6.** Let the scaling matrices take the form (18). Suppose that $\alpha \in \mathcal{C}$, $|\alpha| > 1$, and let $\eta = (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^{n-p-1}$. Then in Case 2 the “optimal” matrix can be chosen as
\begin{equation}
\tilde{D} = \text{diag}(1, \ldots, 1, \eta, \alpha \eta, \ldots, \alpha^{n-p-1} \eta),
\end{equation}
and the “minimal” condition number is
\begin{equation}
\kappa_{\infty}(\tilde{D}R) = \max \left\{ \left( 1 + \sum_{j=2}^{n} |w_j| \right) \left( 1 + \sum_{j=2}^{p} |w_j| \right) + \sum_{j=p+1}^{n} |w_j \alpha^{n-j}|, |\alpha|^{n-p-1} \right\}.
\end{equation}

**Proof.** For any nonsingular matrix $D = \text{diag}(1, \ldots, 1, \delta_{p+1}, \alpha \delta_{p+1}, \ldots, \alpha^{n-p-1} \delta_{p+1})$, we have
\begin{equation}
DR = \begin{bmatrix}
1 & \cdots & w_p & w_{p+1} & \cdots & w_n \\
\vdots & & 1 & \delta_{p+1} & & \\
& & \ddots & \ddots & \ddots & \\
& & & 1 & 1/\delta_{p+1} & \\
& & & & \ddots & \ddots \\
& & & & & 1/(\alpha^{n-p-1} \delta_{p+1})
\end{bmatrix},
\end{equation}
which implies
\begin{equation}
||DR||_{\infty} = \max \left\{ 1 + \sum_{j=2}^{n} |w_j|, |\alpha^{n-p-1} \delta_{p+1}| \right\},
\end{equation}
since $|\alpha| > 1$. On the other hand,
\begin{equation}
R^{-1}D^{-1} = \begin{bmatrix}
1 & -w_2 & \cdots & -w_p & -w_{p+1}/\delta_{p+1} & \cdots & -w_n/(\alpha^{n-p-1} \delta_{p+1}) \\
1 & \ddots & & \ddots & & & \\
& & 1 & 1/\delta_{p+1} & & \\
& & & \ddots & \ddots & \ddots & \\
& & & & 1/(\alpha^{n-p-1} \delta_{p+1}) & & \\
& & & & & & \\
& & & & & &
\end{bmatrix}
\end{equation}
and
\begin{equation}
||R^{-1}D^{-1}||_{\infty} = \max \left\{ 1 + \sum_{j=2}^{p} |w_j| + \sum_{j=p+1}^{n} \left| \frac{w_j}{\alpha^{j-p-1} \delta_{p+1}} \right|, \frac{1}{|\delta_{p+1}|} \right\}.
\end{equation}

(a) If $|\delta_{p+1}| \leq (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^{n-p-1}$, then
\begin{equation}
\kappa_{\infty}(DR) = \max \left\{ \left( 1 + \sum_{j=2}^{n} |w_j| \right) \left[ 1 + \sum_{j=2}^{p} |w_j| + \frac{1}{|\delta_{p+1}|} \sum_{j=p+1}^{n} \left| \frac{w_j}{\alpha^{j-p-1}} \right| \right], \left( 1 + \sum_{j=2}^{n} |w_j| \right) \frac{1}{|\delta_{p+1}|} \right\}.
\end{equation}
In order to minimize the condition number, we have to choose $|\delta_{p+1}|$ as large as possible. Since $|\delta_{p+1}| \leq (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-p-1}$, we can choose $|\delta_{p+1}| = (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-p-1}$. 

(b) If $|\delta_{p+1}| \geq (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-p-1}$, then

$$
\kappa_{\infty}(DR) = \max \left\{ |\alpha|^{n-p-1}|\delta_{p+1}| \left( 1 + \frac{p}{\sum_{j=2}^{n} |w_j|} \right) + \sum_{j=p+1}^{n} |w_j|\alpha^{n-j}, \quad |\alpha|^{n-p-1} \right\}.
$$

In order to minimize the condition number, it is desirable to choose $|\delta_{p+1}|$ as small as possible. Since $|\delta_{p+1}| \geq (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-p-1}$, we can choose $|\delta_{p+1}| = (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-p-1} = \eta$. It is easy to see that (20) is a direct result of choosing $|\delta_{p+1}| = (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-p-1}$ in (21) or (22). 

\[ \text{Case 3}. \]

$$
J_{n_{i}}(\alpha_{n_{i}}) = \begin{bmatrix}
\alpha n_{i} & 0 \\
\alpha n_{i} & 1 \\
\vdots & \ddots \\
\alpha n_{i} & 1
\end{bmatrix} \in \mathbb{C}^{n_{i} \times n_{i}}, \quad i = 1, 2, \ldots, k.
$$

Let $D = \text{diag}(I_{p}, D_{1}, \ldots, D_{k})$, where $I_{p}$ is the $p \times p$ identity matrix and $D_{i} = \text{diag}(d_{i_{1}}, d_{i_{2}}, \ldots, d_{i_{n_{i}}}) \in \mathbb{C}^{n_{i} \times n_{i}}, \quad i = 1, 2, \ldots, k,$ are nonsingular matrices. Similar to the proof in Case 2, we have

$$
D \cdot J(\alpha) = \tilde{J} \cdot D \Leftrightarrow D_{i} \cdot J_{n_{i}}(\alpha_{n_{i}}) = J_{n_{i}}(\alpha_{n_{i}}) \cdot D_{i} \Leftrightarrow d_{i_{1}} = \alpha d_{i_{2}}, \ldots, d_{i_{n_{i}}} = \alpha d_{i_{n_{i}-1}},
$$

so $D$ takes the form

$$
(23) \quad D = I_{p} \oplus \text{diag}(1, d_{i_{1}}, \ldots, \alpha^{n_{1}-2}d_{i_{n_{1}}}) \oplus \cdots \oplus \text{diag}(1, d_{k_{1}}, \ldots, \alpha^{n_{k}-2}d_{k_{n_{k}}}).
$$

For simplicity, we consider $D_{i}$ of the form $D_{i} = \text{diag}(1, \delta, \alpha \delta, \ldots, \alpha^{n_{i}-1} \delta), \quad i = 1, 2, \ldots, k$. Consequently,

$$
(24) \quad D = I_{p} \oplus \text{diag}(1, \delta, \ldots, \alpha^{n_{1}-2} \delta) \oplus \cdots \oplus \text{diag}(1, \delta, \ldots, \alpha^{n_{k}-2} \delta), \quad \delta \neq 0.
$$

Partition the first row of $R^{-1}$ conformably with $D$

$$
[1, -w_{2}, \ldots, -w_{n}] = [U_{p}, V_{1}, \ldots, V_{k}],
$$

where $U_{p} = [1, -w_{2}, \ldots, -w_{p}], \quad V_{i} = [\tilde{w}_{i_{1}}, \ldots, \tilde{w}_{i_{n_{i}}}] \in \mathbb{C}^{n_{i}}$, and $\{\tilde{w}_{i_{1}}, \ldots, \tilde{w}_{i_{n_{i}}}\} \subset \{-w_{p+1}, \ldots, -w_{n}\}, \quad i = 1, 2, \ldots, k$. We have the following theorem.

**Theorem 7.** Let the scaling matrix take the form (24). Suppose that $\alpha \in \mathbb{C}, \quad |\alpha| > 1,$ and denote $n_{i} = \max_{1 \leq i \leq k} \{n_{i}\}$ and $\mu = (1 + \sum_{j=2}^{n} |w_j|) / |\alpha|^{n-2}$. Then in Case 3 the “optimal” matrix can be chosen as

$$
(25) \quad \tilde{D} = I_{p} \oplus \text{diag}(1, \mu, \alpha \mu, \ldots, \alpha^{n_{1}-1} \mu) \oplus \cdots \oplus \text{diag}(1, \mu, \alpha \mu, \ldots, \alpha^{n_{k}-1} \mu),
$$
and the “minimal” condition number is

\[(26)\]

\[\kappa_\infty(\tilde{D}R) = \max \left\{ \left( 1 + \sum_{j=2}^n |w_j| \right) \left( 1 + \sum_{j=2}^p |w_j| + \sum_{i=1}^k |\tilde{w}_i| \right) + |\alpha|^{n_q-2} \Delta, \ |\alpha|^{n_q-2} \right\}, \]

where \( \Delta = \sum_{i=1}^k \left( |\tilde{w}_{i2}| + |\tilde{w}_{i3}|/|\alpha| + \cdots + |\tilde{w}_{in_i}|/|\alpha|^{n_i-2} \right) \).

**Proof.** For any matrix \( D \) that takes the form (24), we have

\[
DR = \begin{bmatrix}
1 & w_2 & \cdots & w_p & \cdots & \cdots & w_n \\
1 & & & & & & \\
& \ddots & & & & & \\
& & 1 & D_1 & & & \\
& & & \ddots & & & \\
& & & & \ddots & & \\
& & & & & \ddots & \\
& & & & & & 1
\end{bmatrix}
\]

and

\[
\|DR\|_\infty = \max \left\{ 1 + \sum_{i=2}^n |w_i|, \ |\alpha|^{n_q-2} \right\},
\]

where \( n_q = \max_{1 \leq i \leq k} \{ n_i \} \). It is easy to verify that

\[
R^{-1}D^{-1} = \begin{bmatrix}
1 & -w_2 & \cdots & -w_p & V_1D_1^{-1} & \cdots & V_kD_k^{-1} \\
1 & & & & & & \\
& \ddots & & & & & \\
& & 1 & D_1^{-1} & & & \\
& & & \ddots & & & \\
& & & & \ddots & & \\
& & & & & \ddots & \\
& & & & & & 1
\end{bmatrix},
\]

so we get

\[
\|R^{-1}D^{-1}\|_\infty = \max \left\{ \|U_p\|_1 + \sum_{i=1}^k \|V_iD_i^{-1}\|_1, \ \frac{1}{|\delta|} \right\}.
\]

Note that

\[V_iD_i^{-1} = [\tilde{w}_{i1}, \tilde{w}_{i2}/\delta, \ldots, \tilde{w}_{in_i}/(\alpha^{n_i-2}\delta)]\]

and

\[\|V_iD_i^{-1}\|_1 = ||\tilde{w}_{i1}| + \frac{1}{|\delta|} (|\tilde{w}_{i2}| + |\tilde{w}_{i3}|/|\alpha| + \cdots + |\tilde{w}_{in_i}|/|\alpha|^{n_i-2}), \quad i = 1, 2, \ldots, k.\]

Therefore,

\[\sum_{i=1}^k \|V_iD_i^{-1}\|_1 = \sum_{i=1}^k ||\tilde{w}_{i1}| + \frac{1}{|\delta|} \sum_{i=1}^k (|\tilde{w}_{i2}| + |\tilde{w}_{i3}|/|\alpha| + \cdots + |\tilde{w}_{in_i}|/|\alpha|^{n_i-2}).\]
In order to minimize the condition number, it is desirable to choose \( \delta \) as large as possible. Since \( |\delta| \leq (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^n - 2 \), we can choose \( |\delta| = (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^n - 2 = \mu \).

(b) If \( |\delta| \geq (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^n - 2 \), then

\[
\kappa_\infty(D\bar{R}) = \max \left\{ |\alpha^{n-2} \delta| \left( ||U_p||_1 + \sum_{i=1}^{k} ||V_i D_i^{-1}||_1 \right), |\alpha|^n - 2 \right\}.
\]

In order to minimize the condition number, it is desirable to choose \( |\delta| \) as large as possible. Since \( |\delta| \geq (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^n - 2 \), we can choose \( |\delta| = (1 + \sum_{j=2}^{n} |w_j|)/|\alpha|^n - 2 = \mu \), and (26) is a direct conclusion from (27) or (28).

Remark. In [9], Serra-Capizzano proposed that the “optimal” diagonal matrix of minimal conditioning can be chosen as

\[
\bar{D} = I_p \oplus \text{diag}(1, \alpha, \ldots, \alpha^{n-1}) \oplus \cdots \oplus \text{diag}(1, \alpha, \ldots, \alpha^{n_k-1}),
\]

which is obviously different from ours since the diagonal elements of the two matrices are different; see (25). Moreover, it seems that our choice is more general.

Acknowledgments. The authors would like to express their sincere thanks to Prof. R. Horn and Prof. S. Serra-Capizzano for their preprints [6, 9] and for many stimulating discussions, and to Prof. D. O’Leary for her insightful suggestions on the previous version of the manuscript.

REFERENCES