AN OPTIMUM ITERATION FOR THE MATRIX POLAR DECOMPOSITION*

A. A. DUBRULLE†

Abstract. It is shown that an acceleration parameter derived from the Frobenius norm makes Newton's iteration for the computation of the polar decomposition optimal and monotonic in norm. A simple machine-independent stopping criterion ensues. These features are extended to Gander's formulas for full-rank rectangular matrices.

Key words. Matrix polar decomposition, Newton iteration.

AMS subject classifications. 65F30, 65F35.

1. Introduction. The polar decomposition of a nonsingular matrix $A \in \mathbf{R}^{n\times n}$ is defined by

$$A = WM, \qquad W, M \in \mathbf{R}^{n\times n}, \quad \begin{cases} W^T W = I, & \\ M = M^T, & x^T M x > 0 \quad \forall \quad x \neq 0. \end{cases}$$

In [3], Higham describes a simple algorithm for the iterative computation of W based on Newton's method applied to the identity $W^T W = I$,

$$\begin{aligned} W^{(0)} &= A, \\ W^{(k+1)} &= \frac{1}{2}\left(\gamma^{(k)} W^{(k)} + \frac{1}{\gamma^{(k)}}[W^{(k)}]^{-T}\right), \qquad k = 1, 2, \ldots, \\ W &= W^{(\infty)}, \end{aligned}$$

where $\gamma^{(k)}$ is an acceleration parameter. The iteration preserves the singular vectors of the iterate, and transforms the singular values $\{\sigma_i^{(k)}\}_{i=1}^n$ of $W^{(k)}$ according to

$$\sigma_i^{(k+1)} = \frac{1}{2}\left(\gamma^{(k)}\sigma_i^{(k)} + \frac{1}{\gamma^{(k)}\sigma_i^{(k)}}\right). \tag{1.1}$$

For appropriate values of $\gamma^{(k)}$, the singular values converge to unity from above after the first iteration. The optimum setting of $\gamma^{(k)}$ to

$$\gamma_2^{(k)} = \left(\sigma_{min}^{(k)}\sigma_{max}^{(k)}\right)^{-1/2}$$

minimizes a bound on $\sigma_{max}^{(k)} = \|W^{(k+1)}\|_2$ and produces monotonic convergence for the ℓ_2 norm of the iterates.

Since it is not practical to compute ℓ_2 norms, software implementations must resort to approximations of $\gamma_2^{(k)}$. Higham's substitute,

$$\gamma_H^{(k)} = \left(\frac{\|[W^{(k)}]^{-1}\|_1 \|[W^{(k)}]^{-1}\|_\infty}{\|W^{(k)}\|_1 \|W^{(k)}\|_\infty}\right)^{1/4}, \tag{1.2}$$

† na.dubrulle@na-net.ornl.gov

* Submitted April 15, 1998. Accepted for publication October 19, 1998. Communicated by G. Golub.

minimizes the product of the bounds

$$\|W^{(k+1)}\|_1 \le \frac{1}{2}\left(\gamma^{(k)}\|W^{(k)}\|_1 + \frac{1}{\gamma^{(k)}}\|[W^{(k)}]^{-T}\|_1\right),$$
$$\|W^{(k+1)}\|_\infty \le \frac{1}{2}\left(\gamma^{(k)}\|W^{(k)}\|_\infty + \frac{1}{\gamma^{(k)}}\|[W^{(k)}]^{-T}\|_\infty\right),$$

but does not guarantee monotonic convergence of the product $\|W^{(k)}\|_1\|W^{(k)}\|_\infty$. This iteration generally delivers an IEEE double-precision solution in about ten or fewer steps with the stopping criterion

$$\|W^{(k+1)} - W^{(k)}\|_1 \le \delta\,\|W^{(k)}\|_1, \tag{1.3}$$

where δ is a constant of the order of machine precision. After convergence to the numerical limit $\tilde{W}$, the other factor of the decomposition is computed as

$$\tilde{M} = \frac{1}{2}\left(\tilde{W}^T A + A^T\tilde{W}\right).$$

Generalizations of this algorithm to rectangular and rank-deficient matrices and modifications for performance enhancement are discussed in [1] and [4].

The approximation (1.2) of the optimal acceleration parameter does not substantially affect convergence in practice, but it does not preserve norm monotonicity, a feature especially interesting for iteration control in software implementations. We show next that an alternate choice for $\gamma^{(k)}$ is optimal and produces monotonic convergence in the Frobenius norm.

2. Acceleration in the Frobenius norm. In [5], Kenney and Laub report good experimental results from the replacement of $\gamma_2^{(k)}$ by

$$\gamma_F^{(k)} = \left(\frac{\|[W^{(k)}]^{-1}\|_F}{\|W^{(k)}\|_F}\right)^{1/2}. \tag{2.1}$$

As proved below, it turns out that the corresponding (Frobenius) iteration is optimal with respect to the associated norm and converges monotonically as fast as the true ℓ_2 iteration.

We first consider optimality. By squaring equation (1.1) and summing over i we get

$$\|W^{(k+1)}\|_F^2 = \frac{1}{4}\left(2n + \gamma^{(k)2}\,\|W^{(k)}\|_F^2 + \frac{1}{\gamma^{(k)2}}\,\|[W^{(k)}]^{-1}\|_F^2\right). \tag{2.2}$$

It is easily verified that $\gamma_F^{(k)}$ minimizes $\|W^{(k+1)}\|_F$ such that

$$\|W^{(k+1)}\|_F^2 = \frac{1}{2}\left(n + \|W^{(k)}\|_F\,\|[W^{(k)}]^{-1}\|_F\right), \qquad \|W^{(k+1)}\|_F \le \|W^{(k)}\|_F. \tag{2.3}$$

Monotonicity derives from the first of the above formulas because the singular values of $W^{(k)}$ are not less than unity for $k \ge 1$, and $\|[W^{(k)}]^{-1}\|_F \le \sqrt{n} \le \|W^{(k)}\|_F$. Hence, the ℓ_2 and Frobenius iterations are equivalent in the sense that they are optimal contractions for their associated norms.

To compare rates of convergence, we consider the Frobenius and ℓ_2 iterates of a matrix with singular values not less than unity (this is the case after the first

```
function w=polar(a)
%
% Initialization:
w=a;
limit=(1+eps)*sqrt(size(a,2));
a=inv(w');
g=sqrt(norm(a,'fro')/norm(w,'fro'));
w=0.5*(g*w+(1/g)*a);
f=norm(w,'fro');
pf=inf;
% Iteration:
while (f>limit) & (f<pf)
   pf=f;
   a=inv(w');
   g=sqrt(norm(a,'fro')/f);
   w=0.5*(g*w+(1/g)*a);
   f=norm(w,'fro');
end
return
```

FIG. 2.1. *Iterative computation of the orthogonal factor of the polar decomposition with acceleration in the Frobenius norm.*

iteration in both methods). By the definition of the acceleration parameter, the Frobenius iterate is bounded above in the Frobenius norm by the ℓ_2 iterate, and therefore converges as fast.

The main advantage of the Frobenius acceleration resides in the low-cost computability of the norm, which can be used for efficient and precise iteration control. Monotonic convergence prescribes that the computation should end when the Frobenius norm of the iterate ceases to decrease, that is, when

$$\mathrm{fl}_\varepsilon\left(\|W^{(k+1)}\|_F\right) \geq \mathrm{fl}_\varepsilon\left(\|W^{(k)}\|_F\right), \qquad k > 1, \tag{2.4}$$

where $\mathrm{fl}_\varepsilon(.)$ is floating-point representation in machine precision ε. A backup test,

$$\mathrm{fl}_\varepsilon\left(\|W^{(k+1)}\|_F\right) \leq (1+\varepsilon)\sqrt{n}, \qquad k \geq 0,$$

sharpens control and may save one iteration. Criterion (2.4) is preferable to the negligibility condition (1.3) in two respects: it is machine independent, and it does away with the computation of $\|W^{(k+1)} - W^{(k)}\|_1$, as only $\|W^{(k+1)}\|_F$, a byproduct of the iteration, is needed. The MATLAB implementation of the computation of $\tilde{W}$ is displayed in Figure 2.1. The initialization step (first iteration) takes the singular values to the interval $[1, \infty[$ where the monotonicity and backup tests apply. The inversion of $W^{(k)}$ dominates the computation, and little can be done to reduce this cost, short of using approximations discussed in [4] that are not likely to preserve essential properties of the iteration.

The algorithm is self-correcting in the sense that the norm minimization automatically takes into account the rounding errors inherent in each iterate. After the first iteration, the norm reduction of the iterates outweighs the forward bound of the rounding errors, and monotonicity is maintained. Self-correction would not take place if we were using the economical formula (2.3) to evaluate $\|W^{(k+1)}\|_F$, with the possibility of a premature termination of the iteration. The effects of rounding errors dwindle in the course of the computation because the Frobenius condition number $\kappa_F(W^{(k)})$ monotonically decreases at the same rate as the norm, as shown by the following identity derived from equation (2.3):

$$\kappa_F(W^{(k)}) = 2\,(\|W^{(k+1)}\|_F + \sqrt{n})\,(\|W^{(k+1)}\|_F - \sqrt{n}).$$

Numerical experiments comparing Higham's and Frobenius accelerations were performed for a wide variety of matrices with assigned singular values, including those cited in [3] and [5]. With the setting $\delta = n\varepsilon$, the former was generally less economical by one step, because the distance between two successive iterates in test (1.3) substantially overestimates the distance to the limit. Both algorithms delivered results accurate to machine precision as measured by $\|I - \tilde{W}^T\tilde{W}\|$ and $\|\tilde{W}^T A - A^T\tilde{W}\|/\|\tilde{W}^T A\|$.

3. Application to full-rank rectangular matrices. The polar decomposition of a full-rank rectangular matrix $A \in \mathbf{R}^{m\times n}$, is defined by

$$A = \begin{cases} WM, & W \in \mathbf{R}^{m\times n}, \quad W^TW = I, \quad M \in \mathbf{R}^{n\times n}, \quad m > n, \\ MW, & W \in \mathbf{R}^{m\times n}, \quad WW^T = I, \quad M \in \mathbf{R}^{m\times m}, \quad m < n. \end{cases}$$

Since each of these definitions derives from the other by substitution of A^T for A, we shall restrict our discussion to the case $m > n$.

Gander [1] generalizes Newton's iteration to full-rank rectangular matrices in $\mathbf{R}^{m\times n}$ as follows:

$$W^{(k+1)} = \frac{1}{2}\,W^{(k)}\left[\gamma^{(k)} I + \frac{1}{\gamma^{(k)}}\left(W^{(k)T}W^{(k)}\right)^{-1}\right], \qquad m > n. \tag{3.1}$$

A singular-value factorization of the iterate,

$$W^{(k)} = U^{(k)}\Sigma^{(k)}V^{(k)T}, \qquad U^{(k)} \in \mathbf{R}^{m\times n}, \qquad \Sigma^{(k)}, V^{(k)} \in \mathbf{R}^{n\times n},$$

yields the same singular-value relation (1.1) as in the case $m = n$. It follows that all derivations concerning the Frobenius acceleration also apply to Gander's formulas. Squaring of equation (1.1) and summation over i produce the expression of $\|W^{(k+1)}\|_F^2$ to be minimized by $\gamma^{(k)}$, but since $W^{(k)}$ does not have an inverse, the identity

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} = \|W^{(k)}(W^{(k)T}W^{(k)})^{-1}\|_F^2$$

provides a substitute for $\|W^{(k)-1}\|_F^2$ in equation (2.2). The optimal value of the acceleration parameter ensues:

$$\gamma_F^{(k)} = \left(\frac{\|W^{(k)}(W^{(k)T}W^{(k)})^{-1}\|_F}{\|W^{(k)}\|_F}\right)^{1/2}.$$

For efficiency in algorithm implementations, this expression suggests the modification of the iteration formula (3.1) to

$$W^{(k+1)} = \frac{1}{2}\left[\gamma^{(k)}W^{(k)} + \frac{1}{\gamma^{(k)}}W^{(k)}\left(W^{(k)T}W^{(k)}\right)^{-1}\right], \qquad m > n.$$

Note that Gander's approach requires the inversion of a matrix whose condition number is the square of that of $W^{(k)}$, which could be numerically harmful in the initial iterations. An alternate approach avoids this drawback by combining a QR factorization and the polar decomposition of the triangular factor, as outlined below:

$$\left.\begin{aligned} A &= QR \\ R &= \Omega M \end{aligned}\right\} \Rightarrow\ W = Q\Omega, \qquad Q \in \mathbf{R}^{m\times n}, \quad Q^T Q = I.$$

4. Conclusion. The properties of the Frobenius acceleration uncovered here re-establish the optimality of Newton's iteration with ℓ_2 acceleration for a norm computable at low cost. These properties allow for portable and efficient software implementations, where monotonicity advantageously replaces negligibility for precise iteration control.

REFERENCES

[1] W. Gander, *Algorithms for the polar decomposition*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 1102–1115.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.

[3] N. Higham, *Computing the polar decomposition—with applications*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1160–1174.

[4] N. Higham and R. Schreiber, *Fast polar decomposition of an arbitrary matrix*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 648–655.

[5] C. Kenney and A. Laub, *On scaling Newton's method for the polar decomposition and the matrix sign function*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 688–706.