Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 07–20

COMPOSITE NORMS AND PERFECT CONDITIONING

JOÃO FILIPE QUEIRÓ

ABSTRACT: We study condition numbers for composite norms and analyze the situations where perfect conditioning occurs.

KEYWORDS: composite norm, condition number, absolute norm, distance to singularity, perfect conditioning.
AMS SUBJECT CLASSIFICATION (2000): 15A12, 15A60, 65F35.

1. Introduction

Composite norms are a class of computationally convenient norms for matrices. In this paper we recall the basic properties of such norms, and we propose two possible ways of defining the (inversion) condition number of a square matrix using a composite norm. For each of these ways, we describe completely which matrices are perfectly conditioned, *i.e.* have condition number equal to 1.

2. Notation and preliminaries

$\mathbb{R}^n$ is the space of $n \times 1$ columns over $\mathbb{R}$, sometimes written as n-tuples. A norm φ on $\mathbb{R}^n$ is determined by its unit ball $B_\varphi = \{x : \varphi(x) \leq 1\}$. The dual norm of φ is defined by

$$\varphi^d(y) = \max_{x \neq 0} \frac{|x^T y|}{\varphi(x)}$$

hence clearly $|x^T y| \leq \varphi(x)\varphi^d(y)$. If M is invertible then $(\varphi \circ M)^d = \varphi^d \circ (M^{-1})^T$.

The Hölder norms are $h_p(x) = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$, $1 \leq p < \infty$, $h_\infty(x) = \max |x_i|$. It is well-known that $h_p^d = h_q$ with $q = \dfrac{p}{p-1}$.

Received July 1, 2007. To appear in *Linear and Multilinear Algebra*.

If φ, ψ are norms on $\mathbb{R}^n$ and A is real $n \times n$, then we define

$$S_{\varphi\psi}(A) = \max_{x \neq 0} \frac{\psi(Ax)}{\varphi(x)}$$

which is just the operator norm of $A : (\mathbb{R}^n, \varphi) \longrightarrow (\mathbb{R}^n, \psi)$. Simple properties of these norms include $S_{\varphi\psi}(A) = S_{\psi^d\varphi^d}(A^T)$, $S_{\varphi\psi}(xy^T) = \psi(x)\varphi^d(y)$, $S_{\varphi\psi}(BA) \leq S_{\nu\psi}(B)S_{\varphi\nu}(A)$, and, for invertible M, $S_{\varphi,\psi\circ M} = S_{\varphi\psi} \circ M$.

Well-known and simple examples are

$$S_{h_1 h_\infty}(A) = \max_{i,j} |a_{ij}|\,,$$

$$S_{h_1 h_1}(A) = \max_{j} \sum_i |a_{ij}|\,,$$

and, by duality,

$$S_{h_\infty h_\infty}(A) = \max_{i} \sum_j |a_{ij}|\,.$$

But the computation of $S_{h_\infty h_1}$ is NP-hard [7]. In general, the effective computation of operator norms of matrices is not easy. For an example of work on this subject, see [1]. An algorithm for finding the $S_{h_p h_p}$ norm of a matrix can be seen in [4].

The (inversion) condition number of a nonsingular square matrix A using the norm $S_{\varphi\psi}$ is defined as

$$\kappa_{\varphi\psi}(A) = S_{\varphi\psi}(A) \cdot S_{\psi\varphi}(A^{-1})\,.$$

Then, as is well-known, if x is the solution of $Ax = b$ and x' is the solution of the perturbed system $Ax = b'$, we have that

$$\frac{1}{\kappa_{\varphi\psi}(A)} \cdot \frac{\psi(b - b')}{\psi(b)} \leq \frac{\varphi(x - x')}{\varphi(x)} \leq \kappa_{\varphi\psi}(A) \cdot \frac{\psi(b - b')}{\psi(b)}\,. \tag{1}$$

If the matrix is perturbed and x' is the solution of $A'x = b$, then

$$\frac{\varphi(x - x')}{\varphi(x')} \leq \kappa_{\varphi\psi}(A) \cdot \frac{S_{\varphi\psi}(A - A')}{S_{\varphi\psi}(A)}\,. \tag{2}$$

Finally, concerning inversion, if A' is a perturbation of A (close to A), we have

$$\frac{S_{\psi\varphi}(A^{-1} - A'^{-1})}{S_{\psi\varphi}(A^{-1})} \leq \frac{\kappa_{\varphi\psi}(A)}{1 - \kappa_{\varphi\psi}(A) \cdot \frac{S_{\varphi\psi}(A-A')}{S_{\varphi\psi}(A)}} \cdot \frac{S_{\varphi\psi}(A - A')}{S_{\varphi\psi}(A)} \,. \tag{3}$$

It is also well-known that, if we denote by $\delta_{\varphi\psi}(A)$ the $S_{\varphi\psi}$-distance of A to the set of singular matrices, we have

$$\kappa_{\varphi\psi}(A) = \frac{S_{\varphi\psi}(A)}{\delta_{\varphi\psi}(A)} \,.$$

Trivially $\kappa_{\varphi\psi}(A) \geq 1$, and a matrix A is perfectly conditioned in the sense of the norm $S_{\varphi\psi}$, *i.e.* $\kappa_{\varphi\psi}(A) = 1$, if and only if $\psi \circ A = \lambda\, \varphi$ for some positive constant λ. If $\psi = \varphi$, this means A is a scalar multiple of a φ-isometry. In the special case $\varphi = h_2$, we essentially get the orthogonal matrices.

3. Composite norms

Recall first that a norm φ on $\mathbb{R}^n$ is absolute if $\varphi(x_1, \ldots, x_n) = \varphi(|x_1|, \ldots, |x_n|)$ for any vector $x = (x_1, \ldots, x_n)$. It is easy to see that φ is absolute if and only if φ^d is absolute. Also [2] φ is absolute if and only if it is monotone, *i.e.* $|x_i| \leq |y_i| \;\;\Rightarrow\;\; \varphi(x_1, \ldots, x_n) \leq \varphi(y_1, \ldots, y_n)$.

We proceed to define composite norms, restricting ourselves to square matrices.

Definition. *Let φ and ψ be norms on $\mathbb{R}^n$, with φ absolute, and let A $n \times n$ have columns $v_1, \ldots, v_n$. The $\varphi\psi$-composite norm of A is*

$$\|A\|_{\varphi\psi} = \varphi(\psi(v_1), \ldots, \psi(v_n))$$

It is easy to see that this actually defines a norm on the space of $n \times n$ matrices, using the fact that φ is monotone to prove the triangle inequality. It is clear that these norms are easy to compute for any matrix.

With the composite norm notation, we can rewrite the examples above as $S_{h_1 h_\infty}(A) = \|A\|_{h_\infty h_\infty}$, $S_{h_1 h_1}(A) = \|A\|_{h_\infty h_1}$. Another example is the Frobenius norm $\left(\sum_{i=1}^{n} |a_{ij}|^2\right)^{1/2}$, which is just $\|A\|_{\varphi\psi}$ with $\varphi = \psi = h_2$.

The main references for composite norms are [5, 6]. The basic properties are the following, all with reasonably straightforward proofs:

(1) $\|xy^T\|_{\varphi\psi} = \psi(x)\varphi(y)$

(2) $\|BA\|_{\varphi\nu} \leq S_{\psi\nu}(B)\|A\|_{\varphi\psi}$

(3) $\|A\|_{\varphi\psi} \geq S_{\varphi^d\psi}(A)$

(4) $\|A\|_{h_\infty\psi} = S_{h_1\psi}(A)$

(This property generalizes the expressions seen for $S_{h_1 h_\infty}$ and $S_{h_1 h_1}$.)

(5) If M is invertible then $\|MA\|_{\varphi\psi} = \|A\|_{\varphi,\psi\circ M}$

(6) If D is an invertible diagonal, then $\|AD\|_{\varphi\psi} = \|A\|_{\varphi\circ D,\psi}$

(7) If P is a permutation matrix, then $\|AP\|_{\varphi\psi} = \|A\|_{\varphi\circ P^T,\psi}$

It is interesting to note that property 4 gives essentially the only case in which there is equality in property 3 for arbitrary A and ψ [6]. (See below for a very slight generalization.)

4. Condition numbers

How should we define the condition number using composite norms? If we want to get estimates similar to (1), (2) and (3) using ψ in the space of data and φ in the space of solutions, property 3 of composite norms suggests that we work with $\|\cdot\|_{\varphi^d\psi}$.

The usual theme is that conditioning should be related to distance to singularity (see [8] for an approach to the same question using componentwise comparisons instead of norms).

Therefore, we are interested in the $\|\cdot\|_{\varphi^d\psi}$-distance of a matrix A to the set of singular matrices. Denoting that distance by $\delta^c_{\varphi^d\psi}(A)$, we have, from results in [3, 6], the following:

Theorem 4.1. $\delta^c_{\varphi^d\psi}(A) = \dfrac{1}{S_{\psi\varphi}(A^{-1})}$.

So one possibility to define the $\|\cdot\|_{\varphi^d\psi}$-condition number of A would be to use the expression

$$\frac{\|A\|_{\varphi^d\psi}}{\delta^c_{\varphi^d\psi}(A)} = \|A\|_{\varphi^d\psi} \cdot S_{\psi\varphi}(A^{-1}) \,.$$

We denote this quantity by $\kappa^*_{\varphi\psi}(A)$.

The problem with this definition is that it involves an operator norm. If we wish to restrict ourselves to the computationally convenient composite norms, then the natural definition, again invoking property 3, is

$$\|A\|_{\varphi^d\psi} \cdot \|A^{-1}\|_{\psi^d\varphi}$$

(which requires ψ absolute). We denote this quantity by $\kappa^c_{\varphi\psi}(A)$. It has been used for particular examples of norms, *e.g.* $\varphi = \psi = h_2$.

Trivially, we have

$$\kappa^c_{\varphi\psi}(A) \ \geq \ \kappa^*_{\varphi\psi}(A) \ \geq \ \kappa_{\varphi\psi}(A) \,.$$

Both quantities deserve the name of condition numbers, as the following are easily proved:

Theorem 4.2. *$\kappa^*_{\varphi\psi}(A)$ satisfies estimates similar to (1), (2) and (3), with $S_{\varphi\psi}$ replaced by $\|\cdot\|_{\varphi^d\psi}$.*

Theorem 4.3. *$\kappa^c_{\varphi\psi}(A)$ satisfies estimates similar to (1), (2) and (3), with $S_{\varphi\psi}$ and $S_{\psi\varphi}$ replaced by $\|\cdot\|_{\varphi^d\psi}$ and $\|\cdot\|_{\psi^d\varphi}$ respectively.*

5. Perfect conditioning

In this section we are interested in knowing which matrices are perfectly conditioned for these two notions of condition number. In other words, when does does $\kappa^*_{\varphi\psi} = 1$ or $\kappa^c_{\varphi\psi} = 1$ occur?

The first case is settled by the following result:

Theorem 5.1. $\kappa^*_{\varphi\psi}(A) = 1$ *if and only if* $\psi \circ A = \varphi$ *with* $\varphi = h_1 \circ \Delta\,,\ \Delta$ *diagonal.*

Proof. $(\Rightarrow)$. Since $\kappa^*_{\varphi\psi}(A) \geq \kappa_{\varphi\psi}(A)$, we already know that $\psi \circ A = \lambda\,\varphi$ for some positive constant λ.

$\kappa^*_{\varphi\psi}(A) = 1$ means $\|A\|_{\varphi^d\psi} \cdot S_{\psi\varphi}(A^{-1}) = 1$. The left-hand side is $\geq \|I\|_{\varphi^d\varphi} \geq S_{\varphi\varphi}(I) = 1$. Therefore, $\|I\|_{\varphi^d\varphi} = 1$.

This implies [6] that $\varphi = h_1 \circ \Delta$ for a diagonal Δ: we have

$$\varphi(x) = \varphi(\Sigma x_i e_i) \leq \Sigma|x_i|\varphi(e_i) \leq \varphi(|x|) \cdot \|I\|_{\varphi^d\varphi} = \varphi(x)\,.$$

Therefore, $\varphi(x) = \Sigma|x_i|\varphi(e_i) = h_1(\Delta x)$ with $\Delta = \mathrm{diag}(\varphi(e_i))$. It only remains to include the constant λ in the diagonal matrix Δ.

$(\Leftarrow)$. Suppose $\psi \circ A = \varphi$ with $\varphi = h_1 \circ \Delta$, so $\psi = \varphi \circ A^{-1} = h_1 \circ \Delta A^{-1}$.

We need a slight generalization of property 4, to the effect that, if Δ is diagonal, we have $\|A\|_{h_\infty \circ \Delta^{-1}, \psi} = S_{h_1 \circ \Delta, \psi}(A)$ (the proof is easy; property 4 is just the case $\Delta = I$).

We then have

$$\begin{aligned} \|A\|_{\varphi^d\psi} &= \|A\|_{h_\infty \circ \Delta^{-1},\, h_1 \circ \Delta A^{-1}} \\ &= S_{h_1 \circ \Delta,\, h_1 \circ \Delta A^{-1}}(A) \\ &= \max_{x \neq 0} \frac{h_1 \circ \Delta A^{-1}(Ax)}{h_1 \circ \Delta(x)} = 1\,. \end{aligned}$$

On the other hand,

$$\begin{aligned} S_{\psi\varphi}(A^{-1}) &= S_{h_1 \circ \Delta A^{-1},\, h_1 \circ \Delta}(A^{-1}) \\ &= \max_{x \neq 0} \frac{h_1 \circ \Delta(A^{-1}x)}{h_1 \circ \Delta A^{-1}(x)} = 1\,, \end{aligned}$$

and the conclusion follows. ∎

Another way of presenting this result is that it describes the situations where a matrix A has all its $\|\cdot\|_{\varphi^d\psi}$-approximation numbers equal (where the k-th approximation number of A is defined as the distance of A to the matrices with rank less than k).

What about κ^c? Recall that

$$\kappa^c_{\varphi\psi}(A) = \|A\|_{\varphi^d\psi} \cdot \|A^{-1}\|_{\psi^d\varphi}$$

Since $\kappa^c_{\varphi\psi}(A) \geq \kappa^*_{\varphi\psi}(A)$, we know from Theorem 5.1 that, if $\kappa^c_{\varphi\psi}(A) = 1$, then $\psi \circ A = \varphi$ with $\varphi = h_1 \circ \Delta$, Δ diagonal. Therefore $\psi = h_1 \circ \Delta A^{-1}$, and this norm must be absolute.

$\kappa^c_{\varphi\psi}(A) = 1$ means $\|A\|_{\varphi^d\psi} \cdot \|A^{-1}\|_{\psi^d\varphi} = 1$. Under the above conditions, the first factor is equal to 1, as before. So the restrictions must come from the second factor. Since $(h_1 \circ \Delta A^{-1})^d = h_\infty \circ \Delta^{-1} A^T$, the question becomes: when do we have $h_1 \circ \Delta A^{-1}$ absolute and $\|A^{-1}\|_{h_\infty \circ \Delta^{-1} A^T, h_1 \circ \Delta} = 1$?

We analyze first the question of which (implicitly invertible) matrices M satisfy $h_1 \circ M$ absolute. Trivially all matrices M of the form DP (or PD) have that property, for diagonal D and P a permutation.

It is easily seen that the set of matrices M for which $h_1 \circ M$ is absolute is closed under pre- and post-multiplication by permutation matrices, under post-multiplication by arbitrary diagonal matrices, and under pre-multiplication by diagonal matrices whose diagonal elements have absolute value equal to 1.

Theorem 5.2. *The norm $h_1 \circ M$ is absolute if and only if M has the form*

$$M = P_1 D^1 \left(\bigoplus_1^k \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \bigoplus I_{n-2k} \right) D P_2$$

where P_1 and P_2 are permutations, D is diagonal, and D^1 is a diagonal whose diagonal elements have absolute value equal to 1. Here $0 \leq k \leq \left[\frac{n}{2}\right]$, so in the direct sum it is possible to have no 2×2 blocks (in which case M is of the form DP), or no tailing identity block.

Proof. ($\Leftarrow$) is trivial, as $|x_i + x_j| + |x_i - x_j|$ clearly depends only on $|x_i|$ and $|x_j|$.

Let us prove ($\Rightarrow$). The first step is the remark that

$$B_{h_1 \circ M} = M^{-1} B_{h_1}. \tag{4}$$

Therefore, if $h_1 \circ M$ is absolute, we know that M^{-1} transforms B_{h_1} into a polyhedron with the symmetries of the unit ball of an absolute norm. Denoting by $e_1, \ldots, e_n$ the vectors in the canonical basis of $\mathbb{R}^n$, the vertices of B_{h_1} are $\pm e_1, \ldots, \pm e_n$. From the identity (4) we deduce that the vertices of $B_{h_1 \circ M}$ are $\pm M^{-1}e_1, \dots, \pm M^{-1}e_n$. These $2n$ vectors are the columns of M^{-1} together with the columns of $-M^{-1}$, and as a set they must, by our hypothesis, be invariant under sign changes of the coordinates.

We proceed to show that no column of M^{-1} can have more than two nonzero elements. We argue by contradiction. Suppose M^{-1} has a column with three nonzero elements (the argument for more than three is analogous). By the obvious reductions there is no loss of generality in assuming the column is $(1, 1, 1, 0, \dots, 0)$. So this is a vertex of $B_{h_1 \circ M}$. Since $h_1 \circ M$ is absolute, $(\pm 1, \pm 1, \pm 1, 0, \dots, 0)$ must all be vertices of $B_{h_1 \circ M}$ as well. We have here eight columns, four in M^{-1} and four in $-M^{-1}$. So M^{-1} has four dependent columns, which is impossible.

If every column of M^{-1} has just one nonzero element, then M^{-1} is of the form DP, and so is M.

It remains to analyze the situation where M^{-1} has columns with two nonzero elements. Take one such column. There is no loss of generality in assuming the two elements are consecutive and both equal to 1. So this column $(\dots, 1, 1, \dots)$ is a vertex of $B_{h_1 \circ M}$. As above, it follows all four columns $(\dots, \pm 1, \pm 1, \dots)$ must be vertices of $B_{h_1 \circ M}$ as well. We may assume that, of these, $(\dots, 1, 1, \dots)$ and $(\dots, 1, -1, \dots)$ are columns of M^{-1}, giving rise to a block of the form $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Let us now show that no other nonzero elements occur in the rows of M^{-1} where this block sits. Look at those two positions in any other column. Clearly the nonzero elements (either one or two) in the column cannot all be in those rows, as we would then have three dependent columns. Suppose the column has $0, a$ in those positions ($a, 0$ would be dealt with in the same way), and b elsewhere, say in the next row, with a, b nonzero. But then M^{-1} has to have a column with $0, a, -b$ (or $0, -a, b$) in the corresponding positions. These two columns, together with the columns in the 2×2 block, make up four dependent columns in M^{-1}, an impossibility.

It only remains to put in diagonal position the $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ blocks, as well as the nonzero elements (scaled to 1) occurring alone in columns. M^{-1} is then reduced to the direct sum form described in the theorem, and so is M, as the $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ blocks, apart from a scalar, are self-inverse. The proof is finished. ∎

Geometrically, what this theorem says is that, apart from permutations and diagonal scalings, the only transformations that map B_{h_1} to the unit ball of an absolute norm are $\pi/4$ rotations around the coordinate axes.

Recall that our question was: when do we have $h_1 \circ \Delta A^{-1}$ absolute and $\|A^{-1}\|_{h_\infty \circ \Delta^{-1} A^T, h_1 \circ \Delta} = 1$? We now know the answer to the first part, and we proceed to analyze the second.

Theorem 5.3. *For Δ diagonal, suppose ΔA^{-1} has the form described in the previous theorem:*

$$\Delta A^{-1} = P_1 D^1 \left(\bigoplus_1^k \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \bigoplus I_{n-2k} \right) D P_2 \,.$$

If $\|A^{-1}\|_{h_\infty \circ \Delta^{-1} A^T, h_1 \circ \Delta} = 1$, then $k = 0$, that is, A^{-1}, and therefore A as well, must have the form permutation $\times$ diagonal.

Proof. Denote by Σ the direct sum appearing between parentheses in the expression of ΔA^{-1}. So A has the form

$$A = P_2^T D^{-1} \Sigma^{-1} D^1 P_1^T \Delta \,.$$

Assuming $k > 0$, we now compute $\|A^{-1}\|_{h_\infty \circ \Delta^{-1} A^T, h_1 \circ \Delta}$.

Using properties 5, 6 and 7 of composite norms, as well as the fact that h_∞ and h_1 are absolute and permutation-invariant, we have

$$\begin{aligned} \|A^{-1}\|_{h_\infty \circ \Delta^{-1} A^T, h_1 \circ \Delta} &= \|\Sigma\|_{h_\infty \circ \Sigma^{-1}, h_1} \\ &= h_\infty \circ \Sigma^{-1}(2, \ldots, 2, 1, \ldots, 1) \end{aligned}$$

where the number of 2's is $2k$.

Since

$$\Sigma^{-1} = \bigoplus_1^k \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \bigoplus I_{n-2k} \,,$$

we have $h_\infty \circ \Sigma^{-1}(2, \ldots, 2, 1, \ldots, 1) = \max\{2, 0, 2, 0, \ldots, 2, 0, 1, \ldots, 1\} = 2$, against our hypothesis. So we must have $k = 0$. ■

Since the reciprocal is easily seen to hold, we conclude:

Theorem 5.4. $\kappa^c_{\varphi\psi}(A) = 1$ *if and only if A has the form permutation $\times$ diagonal, $\varphi = h_1 \circ \Delta$ with Δ diagonal, and $\psi = \varphi \circ A^{-1}$.*

Perfect conditioning for composite norms is therefore exceedingly rare.

6. Final remark

Everything carries over to the complex numbers, except Theorem 5.2 (Theorem 5.3 has no point in that case). I conjecture that only complex matrices M of the form permutation $\times$ diagonal satisfy the property that the norm $h_1 \circ M$ is absolute. If this is true, the conclusion obtained in Theorem 5.4 for the real case remains unchanged for the complex case.

Acknowledgement. I thank Hans Schneider for reference [1].

References

[1] J. Albrecht, and P. P. Klein, *On the matrix norm subordinate to the Hölder norm*, Z. Anal. Anwendungen, 18 (1999), pp. 175–182.
[2] F. L. Bauer, J. Stoer, and C. Witzgall, *Absolute and monotonic norms*, Numer. Math., 3 (1961), pp. 257-264.
[3] P. Franck, *Sur la plus courte distance d'une matrice donnée à l'ensemble des matrices singulieres*, C. R. Acad. Sci. Paris, 256 (1963), pp. 3799-3801.
[4] N. Higham, *Accuracy and stability of numerical algorithms*, SIAM, Philadelphia, PA, 2002.
[5] J.-F. Maitre, *Norme composée et norme associée généralisée d'une matrice*, Numer. Math., 10 (1967), pp. 132-141.
[6] J.-F. Maitre, *Sur Certaines Normes et Fonctionelles dans les Espaces de Matrices et d'Opérateurs*, Thèse d'État, Grenoble, 1974.
[7] J. Rohn, *Computing the norm $\|A\|_{\infty,1}$ is NP-hard*, Linear and Multilin. Algebra, 47 (2000), pp. 195-204.
[8] S. Rump, *Ill-conditioned matrices are componentwise near to singularity*, SIAM Review, 41 (1999), pp. 1025-112.

João Filipe Queiró
CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal
E-mail address: jfqueiro@mat.uc.pt
URL: http://www.mat.uc.pt/~jfqueiro