PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 119, Number 2, October 1993

DIAGONALIZATION IN COMPACT LIE ALGEBRAS AND A NEW PROOF OF A THEOREM OF KOSTANT

N. J. WILDBERGER

(Communicated by Jonathan M. Rosenberg)

ABSTRACT. We exhibit a simple algorithmic procedure to show that any element of a compact Lie algebra is conjugate to an element of a fixed maximal abelian subalgebra. An estimate of the convergence of the algorithm is obtained. As an application, we provide a new proof of Kostant's theorem on the projection of orbits onto a maximal abelian subalgebra.

0

Let $M \in M(n, \mathbb{C})$ be a Hermitian matrix and consider the problem of diagonalizing M, that is, finding a unitary $n \times n$ matrix g such that $g^{-1}Mg$ is diagonal. This problem is essentially equivalent to that of finding the eigenvalues and eigenvectors of M. We propose an algorithm for solving this problem which utilizes the Lie algebra structure of $\mathfrak{g}$, the $n \times n$ skew-Hermitian matrices, and the adjoint action of G, the $n \times n$ unitary group, on $\mathfrak{g}$. In fact our method applies generally to any compact connected Lie group G and its Lie algebra $\mathfrak{g}$.

Fix a maximal torus $T \subseteq G$ with Lie algebra $\mathfrak{t} \subseteq \mathfrak{g}$ and let

$$\mathfrak{g} = \mathfrak{t} \oplus \sum_{\alpha \in \Sigma^+} \mathfrak{g}_\alpha \tag{0.1}$$

be the decomposition of $\mathfrak{g}$ into weight spaces under the adjoint action of T. Here Σ^+ is a set of positive roots and each space $\mathfrak{g}_\alpha$ is two-dimensional. Given $Z \in \mathfrak{g}$, we will write

$$Z = Z_0 + \sum_{\alpha \in \Sigma^+} Z_\alpha \tag{0.2}$$

corresponding to (0.1). The idea is then to choose $\alpha \in \Sigma^+$ such that Z_α has maximum norm and then find $g \in G$ such that $\mathrm{Ad}(g)Z$ has no $\mathfrak{g}_\alpha$ component. This turns out to be essentially a problem in $\mathrm{SU}(2)$, which we can solve using only quadratic equations. If $d(Z)$ denotes the distance from Z to the subspace

Received by the editors November 29, 1990 and, in revised form, February 25, 1992.
1991 *Mathematics Subject Classification.* Primary 22E15; Secondary 58F05.
Key words and phrases. Diagonalization, compact Lie algebra, Kostant's theorem.

$\mathfrak{t}$, then we prove the inequality

$$d(\mathrm{Ad}(g)Z) \le \sqrt{(l-1)/l}\, d(Z) \tag{0.3}$$

where l is the number of positive roots.

We then repeat the procedure, getting a sequence $Z = Z^1, Z^2, \ldots$ converging to a point $Z^\infty \in \mathfrak{t}$. Then Z^∞ is the diagonalized form of $Z \in \mathfrak{g}$. The rate of convergence is controlled by (0.3).

As an application, we use the algorithm to provide a direct proof of a theorem of Kostant [1] which states that if $p\colon \mathfrak{g} \to \mathfrak{t}$ is the orthogonal projection and O_X denotes the $\mathrm{Ad}(G)$ orbit through $X \in \mathfrak{t}$, then

$$p(O_X) = \mathrm{conv}\{O_X \cap \mathfrak{t}\} \tag{0.4}$$

where conv denotes convex hull. It is well known that $O_X \cap \mathfrak{t}$ is a finite set of points, the Weyl group orbit of X, so $p(O_X)$ is a convex polytope. Most existing proofs of Kostant's theorem (for example, the approaches of Atiyah [1, 2], Heckman [4], or Guillemin and Sternberg [3] using symplectic geometry) utilize Morse theory. Our proof is direct and conceptually simple.

1

The problem of diagonalization for $G = \mathrm{SU}(2)$ is easy. The group G consists of all matrices of the form

$$g = \begin{vmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{vmatrix} \tag{1.1}$$

where $\alpha, \beta \in \mathbb{C}$ satisfy $|\alpha|^2 + |\beta|^2 = 1$. The Lie algebra $\mathfrak{g} = \mathrm{su}(2)$ consists of all matrices of the form

$$X = \begin{vmatrix} ix & z \\ -\overline{z} & -ix \end{vmatrix} \tag{1.2}$$

where $x \in \mathbb{R}$ and $z \in \mathbb{C}$, and $\mathfrak{t}$ consists of the one-dimensional diagonal subalgebra. The adjoint action is given by conjugation, so that for g and X as in (1.1) and (1.2),

$$\mathrm{Ad}(g^{-1})X = g^{-1}Xg. \tag{1.3}$$

Given $X \in \mathfrak{g}$, the diagonalization problem is to find $g \in G$ such that $g^{-1}Xg$ is diagonal. We may assume $z \neq 0$. The eigenvalues of X are $i\lambda$ and $-i\lambda$ where $\lambda = (x^2 + |z|^2)^{1/2}$ and the corresponding eigenvectors are

$$\begin{vmatrix} z \\ i(\lambda - x) \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} i(\lambda - x) \\ -\overline{z} \end{vmatrix}. \tag{1.4}$$

Both these vectors have length $d = \sqrt{2\lambda(\lambda - x)}$. It follows that if we set

$$g = \frac{1}{d}\begin{vmatrix} z & i(\lambda - x) \\ i(\lambda - x) & \overline{z} \end{vmatrix} \tag{1.5}$$

then $g \in \mathrm{SU}(2)$ and

$$g^{-1}Xg = \begin{vmatrix} i\lambda & 0 \\ 0 & -i\lambda \end{vmatrix}. \tag{1.6}$$

Note that we have used only quadratic equations to obtain g.

2

Let G be a compact, connected, semisimple Lie group with Lie algebra $\mathfrak{g}$. Let $(\ , \ )$ be a G-invariant positive-definite form on $\mathfrak{g}$ and $| \ |$ the associated norm. Let T be a maximal torus with Lie algebra $\mathfrak{t}$. Let $\Sigma \subseteq \mathfrak{t}^*$ be the root system of G with respect to $\mathfrak{t}$; fix an ordering of the roots with Σ^+ the set of positive roots and Δ the set of simple roots. Then under the adjoint action of T, $\mathfrak{g}$ decomposes as an orthogonal direct sum

$$\mathfrak{g} = \mathfrak{t} \oplus \sum_{\alpha \in \Sigma^+} \mathfrak{g}_\alpha \tag{2.1}$$

where $\mathfrak{g}_\alpha$ is a two-dimensional subspace of $\mathfrak{g}$ such that for $X \in \mathfrak{t}$, $\mathrm{ad}(X)$ acts on $\mathfrak{g}_\alpha$ as

$$\alpha(X) \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix} \tag{2.2}$$

with respect to some orthonormal basis of $\mathfrak{g}_\alpha$. For each $\alpha \in \Sigma$, let $k_\alpha \subseteq \mathfrak{t}$ denote the hyperplane

$$k_\alpha = \{X \in \mathfrak{t} | \alpha(X) = 0\}. \tag{2.3}$$

Let $s_\alpha \colon \mathfrak{t} \to \mathfrak{t}$ denote reflection in the hyperplane k_α and let W, the Weyl group, be the finite group generated by the s_α, $\alpha \in \Sigma$.

Let

$$\mathfrak{t}_+ = \{X \in \mathfrak{t} | \alpha(X) \geq 0 \ \forall \alpha \in \Sigma^+\}. \tag{2.4}$$

Then $\mathfrak{t}_+$ is a fundamental chamber for the action of W so that each $X \in \mathfrak{t}$ is W-conjugate to exactly one element of $\mathfrak{t}_+$.

Now let $Z \in \mathfrak{g}$ and consider the orbit O_Z of Z under the adjoint action, i.e.,

$$O_Z = \{\mathrm{Ad}(g)Z | g \in G\}. \tag{2.5}$$

Then it is well known that $O_Z \cap \mathfrak{t}$ is a finite set and in fact consists of exactly one W orbit. It follows that every adjoint orbit O_Z intersects $\mathfrak{t}_+$ in a unique point.

Let

$$V_Z = O_Z \cap \mathfrak{t} \tag{2.6}$$

and let

$$D_Z = \mathrm{conv}(V_Z). \tag{2.7}$$

Since the action of G preserves the form $(\ , \ )$, all points of V_Z have the same norm and so are vertices of the polytope D_Z.

For $\alpha \in \Sigma^+$, denote the centralizer of k_α in $\mathfrak{g}$ by $\mathrm{cent}_{\mathfrak{g}}(k_\alpha)$. That is

$$\mathrm{cent}_{\mathfrak{g}}(k_\alpha) = \{Z \in \mathfrak{g} | Z \cdot X = 0 \ \forall X \in k_\alpha\} \tag{2.8}$$

where we write $[Z, X] = Z \cdot X$.

Lemma 2.1. $\operatorname{cent}_{\mathfrak{g}}(k_\alpha) = \mathfrak{t} \oplus \mathfrak{g}_\alpha$.

Proof. Let $Z \in \operatorname{cent}_{\mathfrak{g}}(k_\alpha)$ and write the decomposition of Z according to (2.1) as

$$Z = Z_0 + \sum_{\beta \in \Sigma^+} Z_\beta \tag{2.9}$$

with $Z_0 \in \mathfrak{t}$ and $Z_\beta \in \mathfrak{g}_\beta \quad \forall \beta \in \Sigma^+$. Then

$$\begin{aligned} Z \cdot X = 0 \quad \forall X \in k_\alpha & \\ \Leftrightarrow \sum_{\beta\in\Sigma^+} X \cdot Z_\beta = 0 \quad \forall X \in k_\alpha & \\ \Leftrightarrow \sum_{\beta\in\Sigma^+} \beta(X) Z'_\beta = 0 \quad \forall X \in k_\alpha & \qquad (2.10) \\ \text{where } Z'_\beta \in \mathfrak{g}_\beta \text{ is nonzero iff } Z_\beta \text{ is nonzero} & \\ \Leftrightarrow Z_\beta = 0 \quad \forall \beta \neq \alpha. \quad \square & \end{aligned}$$

Denote the orthogonal complement in $\mathfrak{t}$ of k_α by $k_\alpha^\perp$. Then $\mathfrak{h}_\alpha = k_\alpha^\perp \oplus \mathfrak{g}_\alpha$ is a three-dimensional subalgebra of $\operatorname{cent}_{\mathfrak{g}}(k_\alpha)$ isomorphic to su(2), and we have the orthogonal decomposition

$$\operatorname{cent}_{\mathfrak{g}}(k_\alpha) = k_\alpha \oplus \mathfrak{h}_\alpha. \tag{2.11}$$

Define

$$\mathfrak{m}_\alpha = \sum_{\substack{\beta\in\Sigma^+ \\ \beta\neq\alpha}} \mathfrak{g}_\beta. \tag{2.12}$$

Then

$$\mathfrak{g} = k_\alpha \oplus \mathfrak{h}_\alpha \oplus \mathfrak{m}_\alpha \tag{2. 13}$$

is an orthogonal decomposition. This decomposition is preserved under the adjoint action of $\mathfrak{h}_\alpha$.

If H_α denotes the connected Lie subgroup of G with Lie algebra $\mathfrak{h}_\alpha$ then by §1 for any $Z \in \mathfrak{h}_\alpha$, we may find $g \in H_\alpha$ such that $\operatorname{Ad}(g)Z = Z' \in k_\alpha^\perp$. Furthermore, if the $k_\alpha^\perp$ component of Z is nonzero, then we may arrange that the $k_\alpha^\perp$ component of Z' lies in the same Weyl chamber (i.e., half-line) as does that of Z. Applying the same reasoning to an arbitrary $Z \in \mathfrak{g}$ gives us the following.

Lemma 2.2. *Let $\alpha \in \Sigma^+$ and $Z \in \mathfrak{g}$. Then we can find $g \in H_\alpha$ such that if $\operatorname{Ad}(g)Z = Z'$ then*

(i) *Z' has no $\mathfrak{g}_\alpha$ component*;
(ii) *the k_α components of Z and Z' are identical*;
(iii) *the $\mathfrak{m}_\alpha$ components of Z and Z' have the same norm*;
(iv) *the $k_\alpha^\perp$ components of Z and Z' are in the same Weyl chamber of $k_\alpha^\perp$.*

We will refer to the process described in the above lemma as "rotating Z about the hyperplane k_α". For $Z \in \mathfrak{g}$ and $\alpha \in \Sigma^+$, define the distance functions

$$d_\alpha(Z) = |Z_\alpha| \tag{2.14}$$

and

$$d(Z) = \left|\sum_{\alpha\in\Sigma^+} Z_\alpha\right| = \left(\sum_{\alpha\in\Sigma^+} d_\alpha(Z)^2\right)^{1/2}. \tag{2.15}$$

The latter is just the distance from Z to $\mathfrak{t}$. The basic algorithm can now be described. Let $Z \in \mathfrak{g}$, and set $Z^1 = Z$. Construct a sequence $Z^1, Z^2, \ldots$ of elements of $\mathfrak{g}$ recursively as follows. Given Z^{n-1}, find $\alpha \in \Sigma^+$ such that $d_\alpha(Z^{n-1})$ is maximum. Use the formulae of §1 and Lemma 2.2 to find $g \in H_\alpha$ such that $\mathrm{Ad}(g)Z^{n-1} = Z^n$ has no $\mathfrak{g}_\alpha$ component and satisfies the other conditions of the lemma. Then if $|\Sigma^+| = l$,

$$\sum_{\beta\in\Sigma^+} d_\beta(Z^{n-1})^2 \le l d_\alpha(Z^{n-1})^2. \tag{2.16}$$

Thus,

$$\begin{aligned} d(Z^n)^2 &= \sum_{\substack{\beta\in\Sigma^+\\ \beta\neq\alpha}} d_\beta(Z^n)^2 = \sum_{\substack{\beta\in\Sigma^+\\ \beta\neq\alpha}} d_\beta(Z^{n-1})^2 \\ &= \sum_{\beta\in\Sigma^+} d_\beta(Z^{n-1})^2 - d_\alpha(Z^{n-1})^2 \\ &\le \left(\frac{l-1}{l}\right) \sum_{\beta\in\Sigma^+} d_\beta(Z^{n-1})^2 = \left(\frac{l-1}{l}\right) d(Z^{n-1})^2. \end{aligned} \tag{2.17}$$

It follows that $d(Z^n) \to 0$ as $n \to \infty$. Set

$$X^n = p(Z^n). \tag{2.18}$$

Then X^{n-1} and X^n differ only in the $k_\alpha^\perp$ direction and

$$|X^n - X^{n-1}| \le d_\alpha(Z^{n-1}). \tag{2.19}$$

By (2.17) this becomes

$$|X^n - X^{n-1}| \le \left(\frac{l-1}{l}\right)^{(n-2)/2} d(Z_1). \tag{2.20}$$

Thus $\{X^n\}$ is a Cauchy sequence and converges to an element $X^\infty \in \mathfrak{t}$ which is also the limit of the sequence $\{Z^n\}$. Since the orbit O_Z is closed, $X^\infty \in O_Z$ so we have "diagonalized" Z by performing an infinite series of rotations about hyperplanes. Furthermore

$$\begin{aligned} |X^n - X^\infty| &\le \sum_{k=n}^{\infty} \left(\frac{l-1}{l}\right)^{(k-1)/2} d(Z_1) \\ &= \frac{\left(\frac{l-1}{l}\right)^{(n-1)/2}}{1-\left(\frac{l-1}{l}\right)^{1/2}} d(Z_1) \le 2l \left(\frac{l-1}{l}\right)^{(n-1)/2} d(Z_1). \end{aligned} \tag{2.21}$$

Each individual rotation is essentially a rotation in one of a finite number of $SU(2)$ inside G. This is clearly an algorithm that could be implemented in a straightforward fashion on a computer.

3

As an application, we use the algorithm to provide a new proof of a theorem of Kostant. We continue with the notation of the previous sections.

Theorem 3.1 (Kostant [5]). *Let* $X \in \mathfrak{t}$. *Then* $p(O_X) = D_X$.

Proof. Let $Z \in O_X$. Write $Z = Z^1$ and use the above algorithm to find a sequence $Z^1, Z^2, \ldots$ converging to $X^\infty \in \mathfrak{t}$ where the sequence of projections $p(Z^n) = X^n$ also converges to X^∞. Note that X^∞ must be W conjugate to X. If we have rotated Z^{n-1} about the hyperplane k_α to obtain Z^n, then as remarked in the previous discussion, X^n differs from X^{n-1} by an element of $k_\alpha^\perp$ and furthermore X^n is further from the hyperplane k_α than X^{n-1} is. Thus X^{n-1} is between X^n and $s_\alpha(X^n)$. It follows that $X^{n-1} \in D_{X^n}$ and so by W-invariance $D_{X^{n-1}} \subseteq D_{X^n}$. Therefore

$$D_{X^1} \subseteq \cdots \subseteq D_{X^n} \subseteq \cdots \subseteq D_{X^\infty} \tag{3.1}$$

and so

$$X^1 \in D_{X^\infty} = D_X. \tag{3.2}$$

But $X^1 = p(Z)$ and $Z \in O_X$ was arbitrary so that

$$p(O_X) \subseteq D_X. \tag{3.3}$$

To show the reverse inclusion, suppose that $X \in \mathfrak{t}_+$ and $Y \in D_X \cap \mathfrak{t}_+$. Consider a particle moving inside $\mathfrak{t}_+$ which begins at Y and always moves along a direction which is a positive multiple of a simple root. It is thus always moving perpendicularly away from one of the walls of $\mathfrak{t}_+$. Suppose whenever it has a choice (i.e., at the initial stage or whenever it reaches one of the walls of $\mathfrak{t}_+$) it chooses a direction in which it can move unimpeded in a straight line the longest. Clearly the particle would eventually approach infinity so in particular after a finite number of steps it will reach the boundary of D_X, say at a point

$$Y' = X - \sum_{\alpha \in \Sigma^+} r_\alpha \alpha \tag{3.4}$$

where $r_\alpha \geq 0$. Here we have identified α with the unique element in $\mathfrak{t}$ such that

$$\alpha(X) = (\alpha, X) \quad \forall X \in \mathfrak{t} \text{ (so that } \alpha \in k_\alpha^\perp).$$

Then clearly after another finite number of steps along simple root directions, the particle can reach X. Using the results of §1, we can choose $X = Z^1, Z^2, \ldots, Z^k = Z$ of $\mathfrak{g}$ such that Z^{j+1} differs from Z^j only by a rotation about the hyperplane k_{α_j}, where X^{j+1} differs from X^j by a multiple of the simple root α_j. Then $Z \in \mathfrak{g}$ is conjugate to X and $p(Z) = Y$ as required. □

Acknowledgment

The author would like to thank Michael Cowling for some useful remarks.

Bibliography

1. M. F. Atiyah, *Convexity and commuting Hamiltonians*, Bull. London Math. Soc. **14** (1982), 1–15.
2. ——, *Angular momentum, convex polyhedra and algebraic geometry*, Proc. Edinburgh Math. Soc. **26** (1983), 121–138.
3. V. Guillemin and S. Sternberg, *Convexity properties of the moment mapping*, Invent. Math. **67** (1982), 491–513.
4. G. J. Heckman, *Projections of orbits and asymptotic behaviour of multiplicities for compact Lie groups*, Invent. Math. **67** (1982), 333–356.
5. B. Kostant, *On convexity, the Weyl group and the Iwasawa decomposition*, Ann. Sci. Ecole Norm. Sup. **6** (1973), 413–455.

School of Mathematics, University of New South Wales, Kensington, New South Wales, 2033, Australia