Communications in Information and Systems
Vol. 2, No. 3, pp. 283-298, December 2002

005

NEWTON'S ALGORITHM IN EUCLIDEAN JORDAN ALGEBRAS, WITH APPLICATIONS TO ROBOTICS*

UWE HELMKE†, SANDRA RICARDO‡, AND SHINTARO YOSHIZAWA†

Abstract. We consider a convex optimization problem on linearly constrained cones in Euclidean Jordan algebras. The problem is solved using a damped Newton algorithm. Quadratic convergence to the global minimum is shown using an explicit step-size selection. Moreover, we prove that the algorithm is a smooth discretization of a Newton flow with Lipschitz continuous derivative.

1. Introduction. It is a great pleasure to contribute this paper to the special issue, honoring John Moore on the occasion of his 60th birthday. I (U.H.) first met John in 1989 during the end of a short visit of mine at the ANU. Communication between us went along easily right from the start, despite quite different backgrounds in mathematics and control. Initial ideas for future research were immediately flying high at fast speed and at the end of the day we had a new Ph.D. student starting to work on balanced realization flows as well as an emerging plan for longer term research collaborations. The excitement and fun in working with John has been steady ever since and added a lot to my own enjoyment of research and life.

The motivation for this paper comes from a problem in robotics that John started to work on a few years ago, i.e. that of force optimization in dextrous hand grasping. Here the main goal is to achieve the minimization of contact forces without violating nonlinear friction cone constraints. In earlier work [2], [8], linear programming techniques have been used, based on piecewise linear approximations of the cone constraints, but these approaches led to ill conditioning problems. Therefore, an alternative approach had to be found that avoids approximations of the constraints. In the pioneering joint paper [1] by John Moore with M. Buss and H. Hashimoto, the task was reformulated as a convex optimization problem on linearly constrained cones of positive definite matrices, thus making contact with the well established area of convex (semi)definite optimization theory. Linearly convergent gradient flow algorithms were applied to find the global minimum and this led to the first real time solution approach. A challenge remained in finding faster computational schemes. This problem has been resolved in subsequent work [6], where a damped Newton algorithm is

*Received on May 6, 2002; accepted for publication on September 10, 2002. Dedicated to John B. Moore, on the occasion of his 60th birthday. The work of S. R. was partially supported by ISR-Coimbra and the Nonlinear Control Network, NCN, while she was visiting the University of Würzburg. The first and third authors were supported by the grant GIF-I-526-034.06/97.

†Department of Mathematics, University of Würzburg, 97074 Würzburg, Germany. E-mail: helmke@mathematik.uni-wuerzburg.de, yosizawa@ism.ac.jp

‡Department of Mathematics, University of Trás-os-Montes e Alto Douro, Apartado 202, 5001-911 Vila Real, Portugal. E-mail: sricardo@utad.pt

proposed that converges quadratically fast to the optimal solution.

The convex function considered in [1], [6] is of the form

$$\Phi(P) = \operatorname{tr}(P) - \log(\det(P)), \qquad (*)$$

where $P = \operatorname{diag}(P_1, \cdots, P_N)$ is a block diagonal positive definite matrix satisfying some affine constraint. The structure of the block matrices corresponds to the different type of friction constraints. A basic observation in [1] is that the friction constraints are equivalent to positive definiteness of block matrices $P = (P_1, \ldots, P_N)$ with

$$P_i := \begin{bmatrix} \mu_i c_{i,1} & 0 & c_{i,2} \\ 0 & \mu_i c_{i,1} & c_{i3} \\ c_{i,2} & c_{i,3} & \mu_i c_{i,1} \end{bmatrix}$$

in the point contact case and

$$P_i := \begin{bmatrix} c_{i,1} & 0 & 0 & \alpha_i c_{i,2} \\ 0 & c_{i,1} & 0 & \alpha_i c_{i,3} \\ 0 & 0 & c_{i,1} & \beta_i c_{i,4} \\ \alpha_i c_{i,2} & \alpha_i c_{i,3} & \beta_i c_{i,4} & c_{i,1} \end{bmatrix}$$

in the soft finger case, respectively. An equivalent and simplified description is given in [6], using 2×2 real or complex Hermitian matrices. It may appear as a surprise that there are such different, but equivalent positivity characterizations of the force constraints. In fact, this motivates our search for a more unified approach.

In this paper we develop such a unified approach within the broader context of convex optimization in Euclidean Jordan algebras. We consider the cost function $(*)$ on the intersection of a symmetric cone with an affine subspace in an arbitrary Euclidean Jordan algebra. Here the second term in the cost function has a natural interpretation as a barrier function for the symmetric cone, while the first one represents the total finger forces applied. The above mentioned positivity descriptions will be shown to be different manifestions of positivity in a Jordan algebra. Extending the analysis of [6], we consider a damped Newton algorithm in a given Jordan algebra and establish, via a suitable step-size selections, quadratic convergence to the optimum. Although our analysis runs completely similar to that of [6], special care is needed when working in Jordan algebras, due to the lack of associativity of products. Therefore, it may appear to be surprising that such a generalization of [6] on Jordan algebra is in fact possible.

The paper is organized as follows. Section 2 summarizes results and concepts on Jordan algebras. In Section 3, the cost function is introduced and elementary properties established. A damped Newton algorithm is shown to be equivalent to a gradient algorithm. In Section 4, the step-size is determined and the main convergence result is proven in Section 5.

2. Preliminaries. In this section we recall some basic definitions and facts on Jordan algebras. For more details see e.g., [3], [9].

Let V be a finite-dimensional, not necessarily associative, real algebra with a unit element e. For any element x in V, let $L(x)$ be the linear map on V, defined by

$$L(x)\,y = x \circ y, \quad y \in V, \tag{1}$$

where $x \circ y$ denotes the product of x and y in V.

Definition 1. *V is called a Jordan algebra if*

$$\begin{aligned} &(i) \quad x \circ y = y \circ x, \\ &(ii) \quad x \circ \left(x^2 \circ y\right) = x^2 \circ (x \circ y) \end{aligned}$$

hold for all $x, y \in V$.

Let V be a Jordan algebra with a unit element e. Given $u \in V$ let $K[u]$ be the vector space in V spanned by $u, u^2 = u \circ u, u^3 = u^2 \circ u, \ldots$. Thus $K[u]$ is an *associative* subalgebra of V. Moreover, if u and v *commute*, i.e.,

$$u \circ (v \circ x) = v \circ (u \circ x)$$

holds for every $x \in V$, then

$$x \circ (y \circ v) = (x \circ y) \circ v \tag{2}$$

for all $x, y \in K[u]$. If $u \circ v = e$ and u, v commute we call v the *inverse* of u, in formulas $u^{-1} = v$. If u^{-1} exists, we define the power u^{-m}, $1 \leq m$, by $u^{-m} := (u^{-1})^m$. The equation (2) then implies

$$u^r \circ (u^s \circ v) = u^{r+s} \circ v$$

for all integers r, s, if u and v commute.

In Jordan algebras, the notion of a trace or determinant of elements can be defined as in matrix algebras. Let $\mathbb{R}[X]$ denote the algebra of polynomials in a single variable X with real coefficients. Given $x \in V$, let

$$\mathbb{R}[x] := \{p(x) : p \in \mathbb{R}[X]\}. \tag{3}$$

Since V is finite-dimensional, there exists for any $x \in V$ a unique smallest integer k such that $e, x, \ldots, x^k$ are linearly dependent. Denote this integer by $m(x)$. Then the rank of V is defined as

$$\operatorname{rank}(V) = r = \max\{m(x) : x \in V\}. \tag{4}$$

An element $x \in V$ is called regular if $m(x) = r$.

The set of regular elements is open and dense in V. For example, in the non-commutative matrix algebra $V = \mathbb{R}^{n\times n}$ with the usual matrix product, the rank is n and regular elements are precisely those matrices with a single Jordan block for each eigenvalue. For any Jordan algebra V, there exist polynomials $a_1, a_2, ..., a_r \in \mathbb{R}[X]$ such that the minimal polynomial of every regular element X is given by

$$f(\lambda; x) = \lambda^r - a_1(x)\lambda^{r-1} + a_2(x)\lambda^{r-2} + \cdots + (-1)^r a_r(x).$$

The polynomials $a_1, a_2, ..., a_r$ are unique and a_j is homogeneous of degree j. The coefficients $a_1(x) = \operatorname{tr}(x)$, $\quad a_r(x) = \det(x)$ are called the *trace* and *determinant* of x, respectively.

Definition 2. *A real Jordan algebra V is called Euclidean, if the symmetric bilinear form on V*

$$\langle x, y\rangle := \operatorname{tr}(x \circ y), \quad x, y \in V \tag{5}$$

is positive definite.

Example 1. *Let $\mathcal{S}(n)$ denote the set of real symmetric $n \times n$- matrices. Set*

$$X \circ Y := \frac{1}{2}(XY + YX), \quad X, Y \in \mathcal{S}(n).$$

Then $\mathcal{S}(n)$ is a Euclidean Jordan algebra. The trace $\operatorname{tr}(X)$ coincides with the usual trace of matrix, similarly for $\det(X)$.

Example 2. *The Euclidean n-space $\mathbb{R}^n$ can be endowed with the Jordan algebra structure defined by the multiplication operation*

$$(x_1, \ldots, x_n) \circ (y_1, \ldots, y_n) := (x_1y_1 + x_2y_n, x_1y_2 + x_3y_n, ..., \sum_{i=2}^{n} x_iy_{i-1} + x_1y_n).$$

For any $x \in \mathbb{R}^n$ consider the linear operator $L(x) : \mathbb{R}^n \to \mathbb{R}^n$ defined by left multiplication $L(x)y := x \circ y$. The matrix representaion of $L(x)$ with respect to the standard basis of $\mathbb{R}^n$ is then

$$L(x) = \begin{bmatrix} x_1 & & 0 & x_2 \\ & \ddots & & \vdots \\ 0 & & x_1 & x_n \\ x_2 & \cdots & x_n & x_1 \end{bmatrix}.$$

Note that $L(x)$ has the same form as the matrices stated in the introduction for $n = 3, 4$. This gives a Jordan algebra interpretation for these matrices. Moreover, $L(x)$ is positive definite if and only if $x_1 > 0, {x_2}^2 + \cdots + {x_n}^2 < {x_1}^2$.

Example 3. *Let $q(X) \in \mathbb{R}[X]$ be a monic polynomial of degree n and*

$$V := \operatorname{Span}\{p(X) | \deg(p) < n\}$$

denote the n-dimensional $\mathbb{R}$-vector space of polynomials of degree less than n. Define a product operation on V by

$$p_1 \circ p_2 := p_1 p_2 \mod q$$

i.e., as the unique remainder of $p_1 p_2$ up to division by q. Then $(V, \circ)$ is a Jordan algebra. It is Euclidean if and only if $q(X)$ has n-distinct real roots. In fact, the trace form defined as

$$\ll X, Y \gg := \mathrm{tr} L(X \circ Y)$$

satisfies

$$(\ll X^i, X^j \gg)_{i,j=0}^{n-1} = H\left(\frac{q'}{q}\right),$$

where $H(\frac{q'}{q})$ is the $n \times n$ Hankel matrix of $\frac{q'}{q}$.

In any Jordan algebra the following identity holds (See Proposition II.4.3 in [3]).

$$\mathrm{tr}((x \circ y) \circ z) = \mathrm{tr}(x \circ (y \circ z)). \tag{6}$$

In particular, the operator $L(x)$ satisfies

$$\mathrm{tr}(L(x)y \circ z) = \mathrm{tr}(L(x)(y \circ z)) = \mathrm{tr}(y \circ L(x)z) \tag{7}$$

for $x, y, z \in V$. This shows the multiplication operator $L(x)$ is selfadjoint with respect to the scalar product in Definition 2 for all $x \in V$. In a Euclidean Jordan algebra the set of squares

$$Q := \{x^2 | x \in V\}$$

has some important properties summarized by the following theorem.

Theorem 1. *Let V be a Euclidean Jordan algebra.*
(i) The cone $Q = \{x^2 | x \in V\}$ is convex, self-dual, and

$$Q = \{x \in V | L(x) \geq 0\}.$$

(ii) The interior Ω of $Q = \{x^2 | x \in V\}$ is a homogeneous symmetric cone, and

$$\Omega = \{x \in V | L(x) > 0\}.$$

Proof. See Theorem III.2.1 in [3]. □

We need the notion of a quadratic representation $P(x) : V \to V$, defined as

$$P(x) := 2L(x)^2 - L(x^2), \quad x \in V. \tag{8}$$

Note that, in a Euclidean Jordan algebra, $P(x)$ is self-adjoint with respect to the canonical inner product, and $P(x)$ is positive for $x \in \Omega$ ([3], P.55). Moreover, if $\mathcal{A}$ denotes an associative algebra, then

$$x \circ y := \frac{1}{2}(xy + yx), \quad x, y \in \mathcal{A} \tag{9}$$

defines a Jordan algebra structure on $\mathcal{A}$ and one has the simplified expression for $P(x) : \mathcal{A} \to \mathcal{A}$ as

$$P(x)y = xyx.$$

Each element $x \in V$ in a Euclidean Jordan algebra with identity element e admits a *spectral decomposition*. To explain this, we need some further notions. An element $c \in V$ is called *idempotent* if $c^2 = c$. Two idempotents c_1 and c_2 are called *orthogonal* if $c_1 \circ c_2 = 0$. We say that the set of idempotents $c_1, \ldots, c_k$ is a *complete system* of orthogonal idempotents if $c_i^2 = c_i, c_i \circ c_j = 0, (i \neq j), c_1 + \cdots + c_k = e$. An idempotent is called *primitive* if it is non-zero and can not be written as the sum of two non-zero idempotents. A *complete system of orthogonal primitive idempotents* is called a *Jordan frame*. We need the following theorem in Section 4, in order to estimate the Newton step-size.

THEOREM 2. *Let V be a Euclidean Jordan algebra with* $\operatorname{rank}(V) = r$. *For $x \in V$ there exists a Jordan frame $c_1, \ldots, c_r$ and real numbers $\lambda_1, \ldots, \lambda_r$ such that*

$$x = \sum_{i=1}^{r} \lambda_i c_i.$$

The numbers λ_i (with their multiplicities) are uniquely determined by x. Furthermore,

$$\det(x) = \prod_{i=1}^{r} \lambda_i, \quad \operatorname{tr}(x) = \sum_{i=1}^{r} \lambda_i.$$

Proof. See [3], Theorem III.1.2. □

3. Cost Function and Properties. Let V be a Euclidean Jordan algebra and

$$\Omega := \{x \in V | L(x) > 0\}$$

the cone of invertible squares. Given arbitrary linear independent elements $a_1, \ldots,$ $a_m \in V$ and real numbers $b_1, \ldots, b_m$, let

$$\mathcal{C} := \{x \in \Omega | \operatorname{tr}(a_j \circ x) = b_j, \quad j = 1, \ldots, m\}. \tag{10}$$

Without loss of generality we assume in the sequel that $a_1, \ldots, a_m$ are orthonormal, i.e.,

$$\operatorname{tr}(a_i \circ a_j) = \delta_{ij}$$

for $i, j = 1, \ldots, m$. Thus $\mathcal{C}$ is the convex intersection of the open cone Ω and m affine hyperplanes in V. Throughout this paper we assume that the feasibility condition $\mathcal{C} \neq \emptyset$ holds. We consider the minimization problem

$$\min_{x \in \mathcal{C}} \Phi\,(x) \tag{11}$$

for the smooth function $\Phi : \mathcal{C} \to \mathbb{R}$

$$\Phi\,(x) = \operatorname{tr}(x) - \log\left(\det(x)\right), \quad x \in \mathcal{C}.$$

To compute the first and second derivative of Φ note that the tangent space of $\mathcal{C}$ at x is the linear subspace

$$T_x\mathcal{C} = \{\xi \in V | \operatorname{tr}(a_j \circ \xi) = 0, j = 1, \ldots, m\}.$$

By the formula $\frac{\partial x^{-1}}{\partial x} = -P(x)^{-1}$, we have

$$D\Phi\,(x)\,(\xi) = \left\langle e - x^{-1}, \xi \right\rangle$$

and

$$D^2\Phi\,(x)\,(\xi, \eta) = \left\langle P\left(x^{-1}\right)\xi, \eta \right\rangle$$

for $\xi, \eta \in V$, where $D^k\Phi\,(x)$ is the k-th derivative of Φ at $x \in \Omega$.

Proposition 1. *The function $\Phi : \mathcal{C} \longrightarrow \mathbb{R}$ is strictly convex with compact sublevel sets.*

The proof of the above proposition follows directly from the proof of Proposition 2.1 in [5]. As a consequence of Proposition 1, there is a unique local and global minimum of Φ. We denote this unique minimum by x^*.

In the sequel we propose a gradient type algorithm of minimization of Φ. and study its convergence properties.

The projected Euclidean gradient in $T_x\mathcal{C}$ with respect to the canonical scalar product is calculated as

$$\nabla\Phi\,(x) = e - x^{-1} - \sum_{i=1}^{m} \gamma_i a_i, \tag{12}$$

where $\gamma_i = \operatorname{tr}(a_i \circ (e - x^{-1}))$.

Consider the Riemannian metric g on $\mathcal{C}$ defined by

$$g(x; \xi, \eta) := \operatorname{tr}\left(P\left(x^{-1}\right)\xi \circ \eta\right), \qquad \xi, \eta \in T_x\mathcal{C}. \tag{13}$$

Note that

$$g(x; \xi, \eta) := D^2\Phi(x)(\xi, \eta) \tag{14}$$

coincides with the restricted Hessian of Φ on $T_x\mathcal{C}$. By inspection, the gradient with respect to this metric is

$$\operatorname{grad}\Phi\left(x\right)=P\left(x\right)\left(e-x^{-1}-\sum_{i=1}^{m}\beta_i a_i\right) \tag{15}$$

with

$$\begin{bmatrix}\beta_1\\ \vdots\\ \beta_m\end{bmatrix}=G^{-1}\begin{bmatrix}\operatorname{tr}(a_1\circ P(x)(e-x^{-1}))\\ \vdots\\ \operatorname{tr}(a_m\circ P(x)(e-x^{-1}))\end{bmatrix} \tag{16}$$

and

$$G_{i,j}=\operatorname{tr}(a_j\circ P(x)a_i),\quad G=(G_{i,j}). \tag{17}$$

Note that the latter matrix is positive definite. Note also that, by the above formulas,

$$g(x;\operatorname{grad}\Phi(x),\xi)=\operatorname{tr}(\nabla\Phi(x)\circ\xi) \tag{18}$$

for all $\xi\in T_x\mathcal{C}$. Let $H_\Phi(x):T_x\mathcal{C}\to T_x\mathcal{C}$ denote the Hessian operator of Φ at x, i.e.,

$$D^2\Phi(x)(\eta,\xi)=\operatorname{tr}(H_\Phi(x)\eta\circ\xi) \tag{19}$$

for $\forall\eta,\xi\in T_x\mathcal{C}$. The Hessian $H_\Phi(x)$ exists uniquely by nondegeneracy of $D^2\Phi(x)$ on $T_x\mathcal{C}$. Thus

$$\operatorname{grad}\Phi(x)=H_\Phi^{-1}(x)\nabla\Phi(x),\quad x\in\mathcal{C}. \tag{20}$$

Following [6] we consider the damped Newton algorithm for minimization of Φ:

$$x_{k+1}=x_k-\alpha_k H_\Phi^{-1}\left(x_k\right)\nabla\Phi\left(x_k\right). \tag{21}$$

The parameter $\alpha_k>0$ is chosen as large as possible, subject to the downhill inequality constraint $\Phi(x_{k+1})\leq\Phi(x_k)$. Via (20), the damped Newton algorithm is simply the gradient algorithm with respect to the Riemannian metric g, i.e.,

$$x_{k+1}=x_k-\alpha_k\operatorname{grad}\Phi\left(x_k\right). \tag{22}$$

In order to numerically implement the damped Newton algorithm, the step-size α_k has to be appropriately chosen. To this end, we consider at each time instant the "downhill" gradient direction $\Delta=-\operatorname{grad}\Phi\left(x\right)$ in the tangent space $T_x\mathcal{C}$. Since Φ is convex, the line search is a convex minimization task.

4. Explicit Step-Size Selection. In order to obtain an effectively implementable step-size with resulting quadratic convergence rate, we have to find a "good" upper bound on admissible values for α. To estimate the step-size, consider the cost function

$$\phi(\alpha) = \operatorname{tr}(x + \alpha\Delta) - \log\det(x + \alpha\Delta), \quad \Delta \in T_x\mathcal{C}. \tag{23}$$

If $x \in \mathcal{C}$, then $x = y^2$ for some $y \in V$. Note that in any Euclidean Jordan algebra the identity holds

$$\det(P(y)x) = \det(y)^2 \det(x).$$

See ([3], Proposition III. 4.2.). Hence,

$$\begin{aligned}\log\det(x + \alpha\Delta) &= \log\det[P(y)\left(e + \alpha P(y)^{-1}\Delta\right)] \\ &= \log\det\left(y^2\right) + \log\det\left(e + \alpha P(y)^{-1}\Delta\right),\end{aligned}$$

and, therefore

$$\phi(\alpha) = \phi(0) + \alpha\operatorname{tr}(\Delta) - \log\det\left(e + \alpha P(y)^{-1}\Delta\right).$$

The derivative of ϕ then is

$$\phi'(\alpha) = \operatorname{tr}(\Delta) - \operatorname{tr}[\left(e + \alpha P(y)^{-1}\Delta\right)^{-1} \circ P(y)^{-1}\Delta]. \tag{24}$$

In order to obtain the second derivative of ϕ, we use the following Lemmas.

Lemma 1. *Let $a \in V$, then for any $\alpha \in \mathbb{R}$ with $\det P(e + \alpha a) \neq 0$,*

$$\begin{aligned}\frac{d}{d\alpha}\left(\operatorname{tr}\left[(e + \alpha a)^{-1} \circ a\right]\right) &= -\operatorname{tr}\left[(e + \alpha a)^{-2} \circ a^2\right] \\ &= -\operatorname{tr}\left[(e + \alpha a)^{-1} \circ a\right]^2.\end{aligned}$$

Proof. We have

$$\begin{aligned}\frac{d}{d\alpha}\left(\operatorname{tr}\left[(e + \alpha a)^{-1} \circ a\right]\right) &= -\operatorname{tr}\left[(P(e + \alpha a)^{-1} a) \circ a\right] \\ &= -\operatorname{tr}(((e + \alpha a)^{-2} \circ a) \circ a) \\ &= -\operatorname{tr}\left[(e + \alpha a)^{-2} \circ a^2\right] \\ &= -\operatorname{tr}\left[(e + \alpha a)^{-1} \circ a\right]^2.\end{aligned}$$

Notice that the third equality results from the formula,

$$\operatorname{tr}\left[P\left(x^{\frac{1}{2}}\right) y\right] = \operatorname{tr}(x \circ y).$$

(See [3],P.32). □

Using this lemma, we obtain

$$\phi''(\alpha) = \left\| \left(e + \alpha P(y)^{-1} \Delta \right)^{-1} \circ P(y)^{-1} \Delta \right\|^2 . \tag{25}$$

For $\Delta = -\mathrm{grad}\Phi(x) \in T_x\mathcal{C}$ then

$$\begin{aligned} \phi'(0) &= \mathrm{tr}(\Delta) - \mathrm{tr}\left(P(y)^{-1} \Delta \right) \\ &= \mathrm{tr}[\left(e - P(y)^{-1} e \right) \circ \Delta] \\ &= \mathrm{tr}[\left(e - x^{-1} \right) \circ \Delta] \\ &= -\mathrm{tr}\left(\nabla\Phi(x) \circ \mathrm{grad}\Phi(x) \right) \\ &= -\langle \nabla\Phi(x), H_\Phi^{-1}(x) \nabla\Phi(x) \rangle < 0, \end{aligned}$$

where the second equality follows from the selfadjointness of $P(y)^{-1}$. Denote

$$\begin{aligned} \lambda_0(x) &= \sqrt{\mathrm{tr}\left(\nabla\Phi(x) \circ \mathrm{grad}\Phi(x) \right)} \\ &= \sqrt{\langle \nabla\Phi(x), H_\Phi^{-1}(x) \nabla\Phi(x) \rangle}. \end{aligned}$$

Clearly, $\lambda_0^2(x) = -\phi'(0)$. Note also that

$$\lambda_0(x) = \sqrt{\mathrm{tr}(P(x^{-1})\mathrm{grad}\Phi(x) \circ \mathrm{grad}\Phi(x))}. \tag{26}$$

This value of $\lambda_0(x)$ will be used in the estimation of the step-size, and is called the Newton decrement.

Since $V = T_x\mathcal{C} \oplus T_x^\perp\mathcal{C}$ and $T_x^\perp\mathcal{C} = \mathrm{Span}\{a_1, \ldots, a_m\}$, we have also, using $P(x^2) = P(x)^2$,

$$-\phi'(0) = \mathrm{tr}\left(P(x)^{-1} \Delta^2 \right) = \left\| P\left(x^{-\frac{1}{2}} \right) \Delta \right\|^2 = \phi''(0), \tag{27}$$

where the second equality follows from the next lemma.

Lemma 2. *For $x \in \mathcal{C}$ and $z \in V$, we have*

$$\mathrm{tr}\left(P\left(x^{-1} \right) z^2 \right) = \mathrm{tr}\left[\left(P\left(x^{-\frac{1}{2}} \right) z \right)^2 \right].$$

Proof. Since $x \in \mathcal{C}, x = y^2$ for some $y \in V$. Then,

$$\begin{aligned} \mathrm{tr}\left(P\left(x^{-1} \right) z^2 \right) &= \mathrm{tr}\left(P\left(x^{-1} \right) z \circ z \right) = \mathrm{tr}\left(P\left(y^{-2} \right) z \circ z \right) \\ &= \mathrm{tr}\left(P\left(y^{-1} \right) P\left(y^{-1} \right) z \circ z \right) = \mathrm{tr}\left[P\left(y^{-1} \right) \left(P\left(y^{-1} \right) z \circ z \right) \right] \\ &= \mathrm{tr}\left[P\left(y^{-1} \right) \left(z \circ P\left(y^{-1} \right) z \right) \right] = \mathrm{tr}\left[P\left(y^{-1} \right) z \circ P\left(y^{-1} \right) z \right] \\ &= \mathrm{tr}\left[\left(P\left(y^{-1} \right) z \right)^2 \right]. \end{aligned}$$

The first, forth and sixth equations result from the symmetry of the operator $P(x)$ with respect to the scalar product, for all $x \in V$, and the third equality comes from the identity $P\left(x^2 \right) = P(x)^2, x \in V$, (See Theorem 4 in [9], P.57). □

This shows that the Newton decrement can be expressed by

$$\lambda_0(x) = \left\| P\left(x^{-\frac{1}{2}}\right)\Delta \right\|. \tag{28}$$

Let $0 \le t$. The second derivative of ϕ then is

$$\phi''(t) = \operatorname{tr}\left[\left(e + tP\left(x^{-\frac{1}{2}}\right)\Delta\right)^{-2} \circ \left(P\left(x^{-\frac{1}{2}}\right)\Delta\right)^2\right]. \tag{29}$$

In order to approximate the exact Newton step-size, we derive a bound on the second derivative $\phi''(t)$. For this, we need a lemma.

LEMMA 3. *Let $e_1, e_2, ..., e_r$ be a Jordan frame of V and let u be an element of V with spectral decomposition $u = \sum_{i=1}^{r} \mu_i e_i$. Let $\mu = \max\{|\mu_i| : i = 1, 2, ..., r\}$. If $e + tu$ is an invertible element for $0 \le t$ and $t\mu < 1$, then*

$$(1 - t\mu)^{-2} u^2 - (e + tu)^{-2} \circ u^2 \in Q.$$

Proof. Using the spectral decomposition,

$$u^2 = \sum_{i=1}^{r} \mu_i^2 e_i, \quad e + tu = \sum_{i=1}^{r} (1 + t\mu_i) e_i, \quad (e + tu)^{-2} = \sum_{i=1}^{r} (1 + t\mu_i)^{-2} e_i,$$

we have

$$\begin{aligned}(1 - t\mu)^{-2} u^2 - (e + tu)^{-2} \circ u^2 &= (1 - t\mu)^{-2} \sum_{i=1}^{r} \mu_i^2 e_i - \sum_{i=1}^{r} (1 + t\mu_i)^{-2} \mu_i^2 e_i \\ &= \sum_{i=1}^{r} \left[\frac{1}{(1 - t\mu)^2} - \frac{1}{(1 + t\mu_i)^2}\right] \mu_i^2 e_i.\end{aligned}$$

Obviously, $(1 - t\mu)^2 < (1 + t\mu_i)^2$. The result follows. □

In any Euclidean Jordan algebra $y - x \in Q$ implies $y = x + z^2$ for suitable $z \in V$. Since $\operatorname{tr}(z^2) \ge 0$, $\operatorname{tr}(y) \ge \operatorname{tr}(x)$. Thus from Lemma 3 we conclude for $u := P(x^{-\frac{1}{2}})\Delta$.

LEMMA 4. *Let $P(x^{-\frac{1}{2}})\Delta = \sum_{i=1}^{r} \lambda_i e_i$ the decomposition into the Jordan frame and $\lambda(x) := \max\{|\lambda_i| | i = 1, ..., r\}$. For $0 \le t < 1/\lambda(x)$, we have*

$$\phi''(t) \leqslant \frac{\left\| P\left(x^{-\frac{1}{2}}\right)\Delta \right\|^2}{(1 - t\lambda)^2}. \tag{30}$$

□

By monotonicity of the right hand side in t,

$$\sup_{0 \leqslant t \leqslant \alpha} \phi''(t) \leqslant \frac{\left\| P\left(x^{-\frac{1}{2}}\right)\Delta \right\|^2}{(1 - \alpha\lambda)^2}. \tag{31}$$

Thus, by the mean value theorem,

$$\begin{aligned}\left|\phi'(\alpha)-\phi'(0)\right| &\leqslant \left(\sup_{0\leqslant t\leqslant\alpha}\phi''(t)\right)\alpha\\ &\leqslant \frac{\left\|P\left(x^{-\frac{1}{2}}\right)\Delta\right\|^{2}\alpha}{(1-\alpha\lambda)^{2}}\\ &\leqslant -\phi'(0),\end{aligned}$$

where the desired last inequality holds only if α is chosen such that

$$\left\|P\left(x^{-\frac{1}{2}}\right)\Delta\right\|^{2}\alpha+(1-\alpha\lambda)^{2}\phi'(0)\leqslant 0.$$

Since $\phi'(0)=-\lambda_0(x)^2=-\|P(x^{-\frac{1}{2}})\Delta\|^2$, we conclude that the smallest positive root of this quadratic polynomial in α is

$$\alpha_0^{**}(x)=\frac{1+2\lambda(x)-\sqrt{1+4\lambda(x)}}{2\lambda^2(x)}.$$

Consider

$$\alpha_0^{*}(x)=\frac{1+2\lambda_0(x)-\sqrt{1+4\lambda_0(x)}}{2\lambda_0^2(x)}. \tag{32}$$

Recalling the equality (28), we have $\lambda_0(x)=\sqrt{\lambda_1^2+\cdots+\lambda_r^2}$ and therefore

$$\lambda(x)=\max_{i=1,2,\ldots,r}|\lambda_i|\leqslant\lambda_0(x). \tag{33}$$

Noting that the function

$$f:[0,\infty)\longrightarrow\mathbb{R},\quad f(x)=\frac{1+2x-\sqrt{1+4x}}{2x^2}, \tag{34}$$

is strictly monotonically decreasing with $f(0)=1$ and $\lim_{x\to\infty}f(x)=0$, we get

$$0<\alpha_0^{*}(x)\leqslant\alpha_0^{**}(x)\leqslant 1. \tag{35}$$

As a consequence of these considerations we have

LEMMA 5. *If x converges to the critical point x^* of Φ, then*

$$\lim_{x\to x^*}\alpha_0^{*}(x)=1.$$

Furthermore, the function $x\longmapsto\alpha_0^{*}(x)$ on $\mathcal{C}$ is continuous.

5. Main Convergence Theorem. To show the desired convergence result, we need the next two lemmas.

LEMMA 6. *Let U be an open subset of a finite-dimensional normed space V, and $\mathcal{F}:U\longrightarrow U$ be a C^1 map such that the derivative $D\mathcal{F}(u)$ is Lipschitz continuous* *at any $u \in U$ and let $u^* \in U$ denote a fixed point of $\mathcal{F}$ with $D\mathcal{F}(u^*) = 0$. Then the recursion*

$$u_{k+1} = \mathcal{F}(u_k)$$

in U is locally quadratically convergent to u^*.

Proof. See [6], for example. □

Lemma 7. *Let $(V, \langle ., . \rangle)$ be Hilbert space and $\lambda : V \longrightarrow \mathbb{R}$, $\lambda(x) = \|h(x)\|$, with $h : V \longrightarrow V$ a smooth map. For all $x \in V$ with $h(x) \neq 0$ then λ is smooth with*

$$\|D\lambda(x)\| \leqslant \|Dh(x)\|.$$

Proof. From $\lambda(x) = \|h(x)\| = \sqrt{\langle h(x), h(x) \rangle}$, we get

$$D\lambda(x)(\xi) = \frac{\langle Dh(x)\xi, h(x)\rangle}{\lambda(x)}.$$

Hence, using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} |D\lambda(x)(\xi)| &= \frac{|\langle Dh(x)\xi, h(x)\rangle|}{\lambda(x)} \\ &\leqslant \frac{\|Dh(x)(\xi)\| \, \|h(x)\|}{\lambda(x)} \\ &= \|Dh(x)(\xi)\|. \end{aligned}$$

Therefore

$$\|D\lambda(x)\| := \sup_{\xi \neq 0} \frac{|D\lambda(x)(\xi)|}{\|\xi\|} \leqslant \sup_{\xi \neq 0} \frac{\|Dh(x)(\xi)\|}{\|\xi\|} = \|Dh(x)\|.$$

□ Our next result is the main result of this paper.

Theorem 3. *For any $x \in \mathcal{C}$, let $\lambda_0(x)$ be the Newton decrement and let*

$$\alpha_0^*(x) = \frac{1 + 2\lambda_0(x) - \sqrt{1 + 4\lambda_0(x)}}{2\lambda_0^2(x)}.$$

For an initial condition $x_0 \in \mathcal{C}$ the algorithm

$$x_{k+1} = x_k - \alpha_0^*(x_k) \operatorname{grad}\Phi(x_k),$$

converges quadratically fast to the unique global minimum $x^* \in \mathcal{C}$ of Φ. Moreover, the function $x \mapsto \alpha_0^*(x)$ is continuous on $\mathcal{C}$ and satisfies

$$\lim_{x \to x^*} \alpha_0^*(x) = 1.$$

Proof. It only remains to prove the quadratic convergence. In order to do so, we show that the map

$$\begin{aligned} &\mathcal{F} : \mathcal{C} \longrightarrow \mathcal{C}, \\ &\mathcal{F}(x) := x - \alpha_0^*(x) \operatorname{grad}\Phi(x) \end{aligned}$$

satisfies the assumptions of the lemma. This then completes the proof.

First note that $\alpha_0^*(x)$ is smooth for any $x \in \mathcal{C}$ with $x \neq x^*$. Thus $\mathcal{F}$ is smooth and hence $D\mathcal{F}$ is Lipschitz continuous at any $x \neq x^*$.

We show that the derivative of $\mathrm{grad}\Phi(x)$ at x^* is the identity transformation. In fact, by differentiating (15) at the critical point, we obtain

$$\begin{aligned} g(x^*; D\mathrm{grad}\Phi(x^*)(\xi), \eta) &= \mathrm{tr}[P(x^*)^{-1}\{\xi - \sum_{i=1}^{m} D\beta_i(x^*)(\xi)P(x^*)a_i\} \circ \eta] \\ &= \mathrm{tr}[P(x^*)^{-1}\xi \circ \eta - \sum_{i=1}^{m} D\beta_i(x^*)(\xi)a_i \circ \eta] \\ &= \mathrm{tr}(P(x^*)^{-1}\xi \circ \eta) - \sum_{i=1}^{m} D\beta_i(x^*)(\xi)\mathrm{tr}(a_i \circ \eta). \end{aligned}$$

From the definition of the tangent space of $\mathcal{C}$, we conclude

$$\begin{aligned} g(x^*; D\mathrm{grad}\Phi(x^*)(\xi), \eta) &= \mathrm{tr}[P(x^*)^{-1}\xi \circ \eta] \\ &= g(x^*; \xi, \eta) \end{aligned}$$

for $\forall \xi, \eta \in T_{x^*}\mathcal{C}$. Thus $D\mathrm{grad}\Phi(x^*)(\xi) = \xi$ for all $\xi \in T_{x^*}\mathcal{C}$, that is, the derivative of $\mathrm{grad}\Phi(x)$ at $x^* \in \mathcal{C}$ is the identity transformation.

Denote $h(x) = -P\left(x^{-\frac{1}{2}}\right) H_\Phi^{-1}(x^*) \nabla\Phi(x)$.

The Newton decrement $\lambda_0(x) = \|h(x)\|$ is Lipschitz continuous since it is the norm of a smooth function. The function α_0^* is the composition of the smooth function f of (34) with λ_0 and therefore is Lipschitz continuous.

From $\lambda_0(x) = \|h(x)\|$ we get $\|D\lambda_0(x)\| \leqslant \|Dh(x)\|$ for all $x \neq x^*$ (by Lemma 7), and because $Dh(x)$ is a smooth function of x, the derivative of λ_0 is locally bounded around x^*. Applying the chain rule to $\alpha_0^*(x) = f(\lambda_0(x))$, we conclude that the same assertion holds for the derivative of α_0^*.

The derivative of $\mathcal{F}$ at any $x \neq x^*$ is

$$I - \alpha_0^*(x)\, D\mathrm{grad}\Phi(x) - D\alpha_0^*(x)\, \mathrm{grad}\Phi(x).$$

The second summand is the product of a Lipschitz continuous function with a smooth function. Therefore it is Lipschitz continuous. The third summand is a product of a locally bounded function and a smooth function vanishing at x^*.Therefore it is also Lipschitz continuous at x^*. Hence we conclude that the derivative of $\mathcal{F}$ is local Lipschitz continuous.

Moreover,

$$\lim_{x \to x^*} D\mathcal{F}(x) = I - \alpha_0^*(x)\, D\mathrm{grad}\Phi(x^*) = 0. \tag{36}$$

The desired result follows from Lemma 6. □

Note that we have also proven the derivative $D\mathcal{F}(x)$ is Lipschitz continuous in x.

REFERENCES

[1] M. Buss, H.Hashimoto, and J.B.Moore, *Dextrous Hand Grasping Force Optimization*, IEEE Transactions on Robotics and Automation, 12:3(1996), pp. 406–418.

[2] F. T. Cheng and D. E. Orin, *Efficient Algorithm for Optimal Force Distribution-the compact -the dual LP method*, IEEE Trans. on Robotics and Automation, 6:2(1990), pp. 178–187.

[3] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Clarendon Press, Oxford, 1994.

[4] L. Faybusovich, *On a matrix generalization of affine-scaling vector fields*, SIAM J. Matrix Anal. Appl., 16:3(1995), pp. 886–897.

[5] L. Faybusovich, *Euclidean Jordan algebras and generalized affine-scaling vector fields*, Preprint.

[6] U. Helmke, K. Hüper, J. B. Moore, *Quadratically Convergent Algorithms for Optimal Dextrous Hand Grasping*, To appear in IEEE Transactions on Robotics and Automation.

[7] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, London, 1993.

[8] J. Kerr and B. Roth, *Analysis of Multi-fingered Hands*, International Journal of Robotics Research, 4:4(1986), pp. 3–17.

[9] M. Koecher, *The Minnesota Notes on Jordan Algebras and their Applications*, Lecture Notes in Math. 1710, Springer-Verlag, Heidelberg, 1999.