Newton's Method for Optimization in Jordan Algebras

Sandra Ricardo *, Uwe Helmke †, and Shintaro Yoshizawa †

May 25, 2002

Abstract

We consider a convex optimization problem on linearly constrained cones in a Euclidean Jordan algebra. The cost function consists of a quadratic cost term plus a penalty function. A damped Newton algorithm is proposed for minimization. Quadratic convergence to the global minimum is shown using an explicit step-size selection.

1 Introduction

Grasping and manipulation of objects plays an important role in robotics, with a strong incentive towards real time implementations. In the area of dextrous hand grasping, positive definite programming yields a satisfactory solution to optimization of finger forces, subject to balancing of external forces and friction cone constraints. In [1], the grasping force optimization task has been formulated as the minimization of the convex function

$$\Phi(X) = \operatorname{tr}(X) - \log(\det X) \tag{1.1}$$

on positive definite matrices satisfying linear equality constraints. This approach led to the first real time solution. In [5], a damped Newton algorithm has been proposed with faster, quadratic convergence to the optimum.

In this paper we propose an extension of this prior work within the more general context of convex optimization in a Jordan algebra. Specifically we consider the optimization of functions

$$\Phi(x) = \operatorname{tr}(p(x)) - \log(\det x) \tag{1.2}$$

*Department of Mathematics, University of Trás-os-Montes e Alto Douro, Apartado 202, 5001-911 Vila Real, Portugal. †Department of Mathematics, University of Würzburg, 97074 Würzburg, Germany. 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 second and third authors were supported by the grant GIF-I-526-034.06/97. Email: sricardo@utad.pt, helmke@mathematik.uni-wuerzburg.de, and yosizawa@ism.ac.jp

on the intersection of a finite number of affine subspaces with the cone of positive elements of a Euclidean Jordan algebra. Here $p(x) = c_0 + c_1 \circ x + \frac{1}{2} c_2 \circ x^2$ denotes an arbitrary quadratic polynomial. A damped Newton algorithm for minimizing (1.2) is proposed, together with an explicit step-size that guarantees quadratic convergence. Since symmetric matrices form a Euclidean Jordan algebra, our results immediately generalize those of [5]. In this short paper no proofs are given. For a complete analysis we refer to [7]. Independent of the applications in robotics, the minimization of cost functions (1.2) can be of interest in other areas as well. For example, the minimization of the unconstrained cost function

$$\Phi(x) = \operatorname{tr}(c \circ x) - \log(\det x) \tag{1.3}$$

on positive elements x is equivalent to solving the linear equation $c \circ x = e$. In the Jordan algebra of symmetric matrices this is just equivalent of solving the Sylvester equation AX+XA = I. Similarly, other interesting equations on matrices or polynomials can also be recast as optimization problems in a Jordan algebra.

2 Cost function and properties

For references on Jordan algebras we refer to [2] and [8]. Let V be a Euclidean Jordan algebra and $\Omega := \{x \in V | x > 0\}$ denote the cone of invertible squares. Similarly, we write $x \geq 0$ to denote the squares $x = y^2$ of V. Given arbitrary linear independent elements $a_1, \ldots, a_m \in V$ and real numbers $b_1, \ldots, b_m$, let

$$\mathcal{C} := \{x \in \Omega | \operatorname{tr}(a_j \circ x) = b_j, \quad j = 1, \ldots, m\}. \tag{2.4}$$

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 $\min_{x \in \mathcal{C}} \Phi(x)$ for the smooth function

$$\Phi(x) = \operatorname{tr}(p(x)) - \log(\det(x)), \quad x \in \mathcal{C}, \tag{2.5}$$

where $p(x) := c_0 + c_1 \circ x + \frac{1}{2} c_2 \circ x^2$ and $c_2 \geq 0$. To compute the first and second derivative of Φ note that the tangent space of $\mathcal{C}$ at x is the linear subspace

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

Then we have

$$D\Phi(x)(\xi) = \left\langle c_1 + c_2 \circ x - x^{-1}, \xi \right\rangle \quad \text{and} \quad D^2\Phi(x)(\xi, \eta) = \left\langle c_2 \circ \xi + P\left(x^{-1}\right)\xi, \eta \right\rangle$$

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

Proposition 2.1. *Let $c_2 \geq 0$. The function $\Phi : \mathcal{C} \longrightarrow \mathbb{R}$ is strictly convex. Φ has compact sublevel sets, provided $c_2 > 0$ or $c_2 = 0$, $c_1 > 0$.*

See also [4]. In the sequel, we will always assume that $c_2 > 0$ or $c_2 = 0$, $c_1 > 0$ holds. As a consequence of Proposition 2.1 , there is a unique local and global minimum of Φ. We denote this unique minimum by x^*.

We propose a gradient type algorithm for minimization of Φ and study its convergence properties; see also [6]. The projected Euclidean gradient in $T_x\mathcal{C}$ with respect to the canonical scalar product is calculated as

$$\nabla\Phi(x) = c_1 + c_2 \circ x - x^{-1} - \sum_{i=1}^{m} \gamma_i a_i, \tag{2.6}$$

where $\gamma_i = \mathrm{tr}(a_i \circ (c_1 + c_2 \circ x - x^{-1}))$.

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

$$g(x;\xi,\eta) := D^2\Phi(x)(\xi,\eta), \qquad \xi,\eta \in T_x\mathcal{C}. \tag{2.7}$$

Note, by strict convexity of Φ, that g is positive definite on each tangent space.

The gradient with respect to the Riemannian g is the uniquely determined vector field $\mathrm{grad}\Phi$ satisfying the identity

$$D^2\Phi(x)(\mathrm{grad}\Phi(x),\xi) = D\Phi(x)\xi, \quad \forall\xi \in T_x. \tag{2.8}$$

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

$$D^2\Phi(x)(\eta,\xi) = \mathrm{tr}(H_\Phi(x)\eta \circ \xi) \tag{2.9}$$

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

$$\mathrm{grad}\Phi(x) = H_\Phi^{-1}(x)\nabla\Phi(x), \quad x \in \mathcal{C}. \tag{2.10}$$

Following [5] we consider the damped Newton algorithm for minimization of Φ:

$$x_{k+1} = x_k - \alpha_k H_\Phi^{-1}(x_k)\nabla\Phi(x_k). \tag{2.11}$$

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 (2.10), the damped Newton algorithm is simply the gradient algorithm with respect to the Riemannian metric g, i.e.,

$$x_{k+1} = x_k - \alpha_k \mathrm{grad}\Phi(x_k). \tag{2.12}$$

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 $\Delta = -\mathrm{grad}\Phi(x)$ in the tangent space $T_x\mathcal{C}$. Since Φ is convex, the line search is a convex minimization task.

3 Step-size selection and main theorem

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. Let $\Delta := -\mathrm{grad}\Phi(x) \in T_x\mathcal{C}$. To estimate the step-size, consider the cost function for $t \geq 0$

$$\phi(t) = \Phi(x + t\Delta) = \mathrm{tr}(p(x + t\Delta)) - \log\det(x + t\Delta). \tag{3.13}$$

Let $P(x) : V \to V$ be the quadratic representation defined as

$$P(x)y := 2x \circ (x \circ y) - x^2 \circ y$$

for $x, y \in V$. Then $P(x)$ is selfadjoint and $P(x) > 0$ if and only if $x > 0$. The second derivative of ϕ is

$$\phi''(t) = \mathrm{tr}(c_2 \circ \Delta^2) + \mathrm{tr}\left[\left(e + tP\left(x^{-\frac{1}{2}}\right)\Delta\right)^{-2} \circ \left(P\left(x^{-\frac{1}{2}}\right)\Delta\right)^2\right]. \tag{3.14}$$

For $\Delta = -\mathrm{grad}\Phi(x) \in T_x\mathcal{C}$ then $\phi'(0) = -\langle \nabla\Phi(x), H_\Phi^{-1}(x)\nabla\Phi(x)\rangle < 0$. Let $\lambda_0(x) = \sqrt{\mathrm{tr}(\nabla\Phi(x) \circ \mathrm{grad}\Phi(x))} = \sqrt{\langle \nabla\Phi(x), H_\Phi^{-1}(x)\nabla\Phi(x)\rangle}$ denote the Newton decrement. Clearly, $\lambda_0^2(x) = -\phi'(0)$. By inspection,

$$-\phi'(0) = \phi''(0). \tag{3.15}$$

We have the following bound on the second derivative $\phi''(t)$

$$\sup_{0 \leqslant t \leqslant \alpha} \phi''(t) \leqslant \mathrm{tr}(c_2 \circ \Delta^2) + \frac{\mathrm{tr}(P(x^{-1})\Delta^2)}{(1 - \alpha\lambda_0^*)^2}, \tag{3.16}$$

where

$$\lambda_0^* := \sqrt{\mathrm{tr}(P(x^{-1})\Delta^2)}. \tag{3.17}$$

Thus, by the mean value theorem,

$$\begin{aligned} |\phi'(\alpha) - \phi'(0)| &\leqslant \left(\sup_{0 \leqslant t \leqslant \alpha} \phi''(t)\right)\alpha \\ &\leqslant \alpha\left[\mathrm{tr}(c_2 \circ \Delta^2) + \frac{\mathrm{tr}(P(x^{-1})\Delta^2)}{(1 - \alpha\lambda_0^*)^2}\right] \leqslant -\phi'(0) = \phi''(0), \end{aligned}$$

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

$$\alpha\left[\mathrm{tr}(c_2 \circ \Delta^2) + \frac{\mathrm{tr}(P(x^{-1})\Delta^2)}{(1 - \alpha\lambda_0^*)^2}\right] \leq \mathrm{tr}(c_2 \circ \Delta^2) + \mathrm{tr}(P\left(x^{-1}\right)\Delta^2). \tag{3.18}$$

Let α_0^* of x denote the smallest positive root of the quadratic polynomial in α

$$\alpha\mathrm{tr}(P(x^{-1})\Delta^2) = (1 - \alpha\lambda_0^*)^2\mathrm{tr}(P(x^{-1})\Delta^2). \tag{3.19}$$

Then

$$\alpha_0^*(x) = \frac{1 + 2\lambda_0^*(x) - \sqrt{1 + 4\lambda_0^*(x)}}{2\lambda_0^{*2}(x)}. \tag{3.20}$$

Then we get

$$0 < \alpha_0^*(x) \leqslant 1. \tag{3.21}$$

Since $\alpha_0^*(x)\mathrm{tr}(c_2 \circ \Delta^2) \leq \mathrm{tr}(c_2 \circ \Delta^2)$, we conclude

$$|\phi'(\alpha) - \phi'(0)| \leq -\phi'(0)$$

for all $\alpha \leq \alpha_0^*(x)$. Note that for the Newton decrement λ_0

$$\lambda_0^2(x) = \mathrm{tr}(c_2 \circ \Delta^2) + \mathrm{tr}(P(x^{-1})\Delta^2) \geq (\lambda_0^*)^2.$$

Our main convergence result is:

Theorem 3.1. *Assume $c_2 > 0$ or $c_2 = 0, c_1 > 0$. For any $x \in \mathcal{C}$, let $\lambda_0^*(x)$ be defined by (3.17) and let*

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

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

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

converges quadratically fast to the unique global minimum $x^* \in \mathcal{C}$ of

$$\Phi(x) = \mathrm{tr}(c_0 + c_1 \circ x + \frac{1}{2}c_2 \circ x^2) - \log(\det x).$$

References

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

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

[3] L. Faybusovich. On a matrix generalization of affine-scaling vector fields, *SIAM J. Matrix Anal. Appl.*, vol. 16, No. 3, pp. 886-897, July 1995.

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

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

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

[7] U. Helmke, S. Ricardo, and S. Yoshizawa. Newton's Algorithm in Euclidean Jordan Algebras, with Applications to Robotics, To appear in *Communications in Information and Systems*.

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