Matthias Himmelmann

matthias.himmelmann@tu-braunschweig.de

Technische Universität Braunschweig

Riemannian Optimization over Semialgebraic Sets†

1. Constrained Polynomial Optimization

Consider the constrained optimization problem

$$\min_{x \in \mathcal{C}} f(x) \tag{1}$$

for a polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ and a semialgebraic set $\mathcal{C}$ generated by equalities $g_1(x), \dots, g_s(x) = 0$ and inequalities $h_1(x), \dots, h_t(x) \geq 0$ for polynomials $g_i, h_j \in \mathbb{R}[x_1, \dots, x_n]$.

A major obstacle in finding the optimum of a constrained polynomial optimization problem is that both the polynomial p and the constraint set $\mathcal{C}$ are not convex in general. Solution strategies involve reformulating this optimization problem as the maximization of λ such that $f(x) - \lambda \geq 0$ over $\mathcal{C}$, relating polynomial optimization and nonnegativity certificates. Typical relaxations are Sums of Squares[1], leading to a convex optimization problem that can be solved via semidefinite programming, and Sums of Nonnegative Circuits[2]. Such relaxations provide lower bounds on the global optimum of (1).

For finding upper bounds, currently either a highly specialized approach is employed[3] or Newton's method is directly applied. This approach is not guaranteed to converge globally to the closest minimum.

2. Riemannian Geometry

Definition: Let $\mathcal{S} \hookrightarrow \mathcal{M}$ be an embedded Riemannian manifold and $x \in \mathcal{M}$. A *retraction* at x is a smooth map $R_x : T_x\mathcal{M} \to \mathcal{M}$ such that each curve $c(t) = R_x(tv)$ satisfies

$$c(0) = x \text{ and } \dot{c}(0) = v.$$

If additionally $\ddot{c}(0) \in N_x(\mathcal{S}, \mathcal{M})$, it is called 2nd-order.

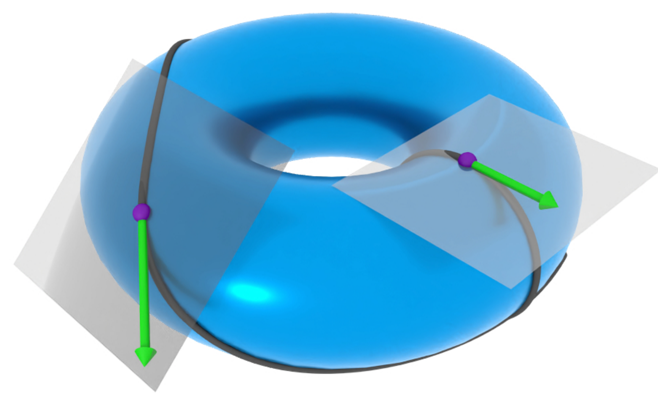


Figure: Two second-order retractions on the torus.

Theorem[4]: Let $\mathcal{M} \subset \mathbb{R}^n$ be a smooth manifold. For any point $p \in \mathcal{M}$, define the relation $R_p \subset T_p\mathcal{M} \times \mathcal{M}$ by

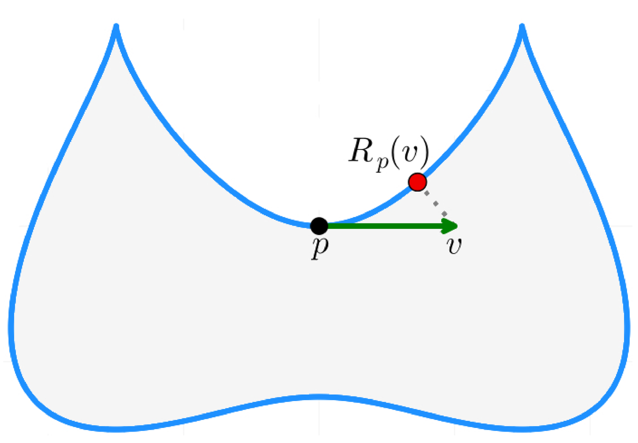
$$R_p = \{(v, u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in \operatorname{argmin}_{y \in \mathcal{M}} |p + v - y|\}.$$

There exists a neighborhood $\mathcal{U}$ of 0 in $T_p\mathcal{M}$ such that R_p defines a local, second-order retraction. I.e., the curve $t \mapsto R_p(tv)$ matches the geodesic at p corresponding to v up to second-order.

3. Euclidean Distance Retraction

The closest point problem: For a smooth $\mathcal{C} = g^{-1}(0) \cup \{x : h_j(x) \geq 0\}$ in $\mathbb{R}^n$ and a fixed point $u \in \mathbb{R}^n$, the closest point problem can be expressed as

$$\min_x \frac{1}{2}|x - u|^2$$
$$\text{s.t. } g_i(x) = 0 \text{ and } h_j(x) \geq 0$$
$$\text{with } i \in \mathcal{E} \text{ and } j \in \mathcal{I}.$$



In terms of *Karush-Kuhn-Tucker conditions*[5], this problem can be reformulated as

$$\mathcal{L}(x, \lambda, \mu; u) = \frac{1}{2}\sum_{i=1}^{n}(x_i - u_i)^2 - \lambda^T g(x) - \mu^T h(x).$$

In $p \in \mathcal{C}$, the set of *active indices* is $\mathcal{A}(p) = \{i : i \in \mathcal{E} \text{ or } h_i(p) = 0\}$. From now on, assume that the set of gradients of active constraints in p is linearly independent (LICQ). The first-order optimality criterion then becomes

$$\nabla_x \mathcal{L}(x, \lambda, \mu; u) = 0$$
$$g_i(x) = 0 \text{ and } h_j(x) \geq 0,$$
$$\mu_j \geq 0 \text{ and } \mu_j \cdot h_j(x) = 0.$$

In particular, the Lagrange mulitpliers corresponding to the inactive inequality constraints are zero. Hence. we can "track" the known solution $x = p$ along a path $u : [0, 1] \to \mathbb{R}^n$ from $u_1 = p$ to $u_0 = p + v$ via the straight-line homotopy

$$H(x, \lambda, \mu; t) = \nabla\mathcal{L}(x, \lambda, \mu; (1 - t)u_0 + t\, u_1)$$

using a predictor-correct scheme known as *homotopy continuation*[6]. It uses an iterative combination of Euler's and Newton's method; whenever $h_j(x) < 0$, we add it to the equality constraints to correct the point to the boundary. Otherwise, h_j does not contribute to this step. Certified curve-tracking methods are applicable[7].

4. Optimality Condition over Semialgebraic Sets

Since the LICQ holds, the *tangent cone* in p is equal to[5]

$$\hat{T}_p\mathcal{C} = \{w : w^T \nabla g_i(p) = 0 \text{ for } i \in \mathcal{E} \text{ or } w^T \nabla h_i(p) \geq 0 \text{ for } j \in \mathcal{A}(p) \cap \mathcal{I}\}.$$

Theorem[5]: Assume for some $p \in \mathcal{C}$ there is a Lagrange multiplier λ, μ such that the KKT conditions are satisfied. If additionally,

$$w^T \nabla^2_{xx}\mathcal{L}(p, \lambda, \mu)\, w > 0 \quad \text{for all nonzero } w \in \hat{T}_p\mathcal{C}$$

with $w^T \nabla h_j(p) = 0$ for all j with $\mu_j > 0$, then p is a strict local minimum.

5. Riemannian Gradient Descent

For a Riemannian manifold $\mathcal{M}$, consider the constrained optimization problem

$$\min_{x \in \mathcal{M}} f(x)$$

with a smooth $f : \mathcal{M} \to \mathbb{R}$. Similar to unconstrained optimization, local methods such as gradient descent can be employed to solve this problem[8]: Given a descent direction $v_k \in \hat{T}_{x_k}\mathcal{M}$ satisfying the conditions from Zoutendijk's theorem[9] and a step size α_k satisfying the Wolfe conditions[9], the retraction is applied repeatedly until f's Riemannian gradient is smaller than a given tolerance $\tau > 0$:

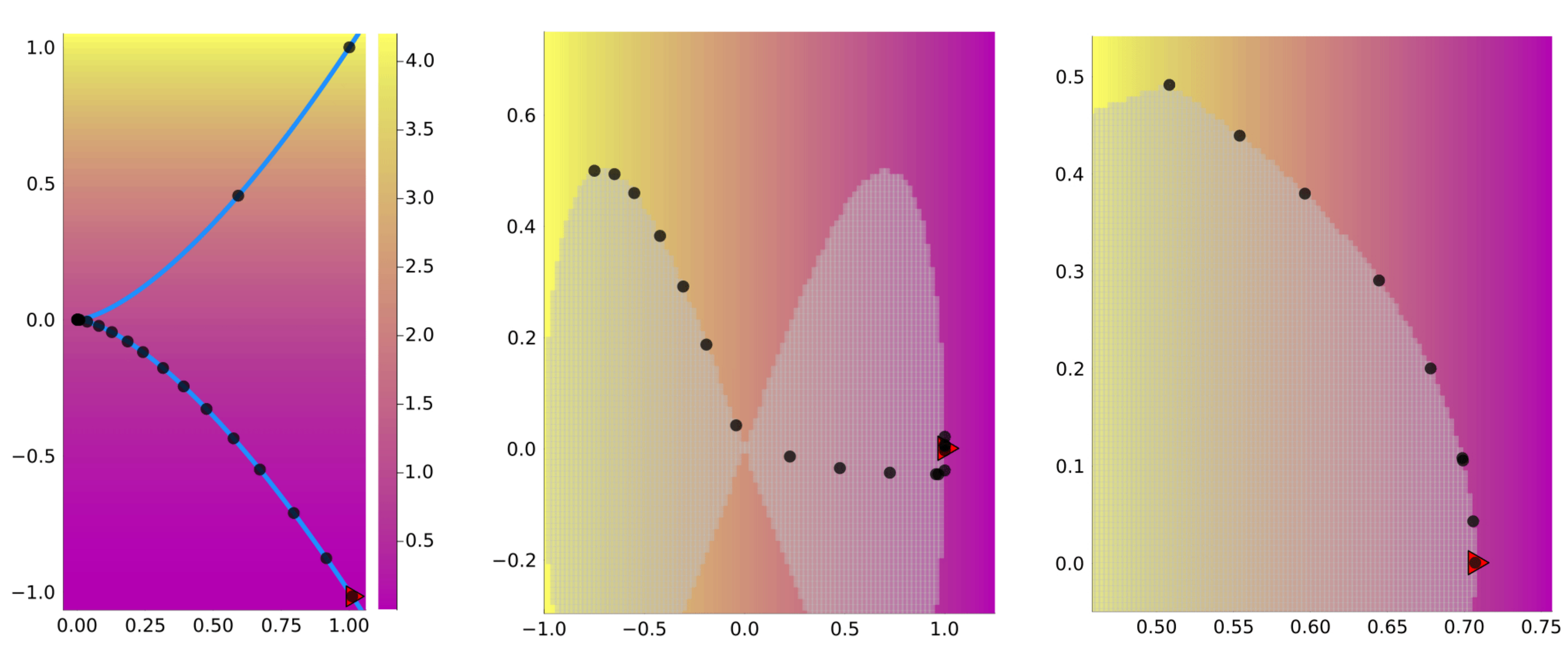
$$x_{k+1} = R_{x_k}(\alpha_k v_k) \text{ until } ||\operatorname{grad} f(x_k)|| < \tau.$$

Here, the Riemannian gradient $\operatorname{grad} f(x)$ is given by the projection of $\nabla f(x)$ onto the tangent cone $\hat{T}_x\mathcal{M}$. In the case of semialgebraic sets $\mathcal{C}$, this is the projection onto the linear subspace generated by the equality and the violated inequality constraints $v^T \nabla h_j(x) < 0$ with $j \in A(x) \cap \mathcal{I}$ for a descent direction $v \in \hat{T}_x\mathcal{C}$.

6. Singularities

In singularities, the dimension of the tangent space may be larger than expected. Hence, there may only be a subset of tangent directions with measure zero that linearizes the local behavior of the constraint set. Given a semialgebraic set $\mathcal{C}$, the set of singular points has a positive codimension and the regular loci of the irreducible components form smooth manifolds[10]. By taking random and small tangent steps weighted by the Riemannian gradient in the singularity, we are able to reliably escape singularities. The optimization routine proceeds if a lower objective function value is identified; otherwise, the singularity is considered optimal.

7. Examples



†This project is joint work with Timo de Wolff.

[1] D. Hilbert: Über die Darstellung definiter Formen als Summe von Formenquadraten. Math. Ann. **32.3** (1888).
[2] S. Iliman, T. de Wolff: Amoebas, Nonnegative Polynomials and SOS Supported on Circuits. R. Math. Sc. **3.1** (2016).
[3] J. B. Lasserre: A New Look at Nonnegativity on Closed Sets and Polynomial Optimization. SIAM J. Opt. **21.3** (2011).
[4] P.-A. Absil and J. Malick. Projection-like Retractions on Matrix Manifolds. SIAM J. Opt. **22.1** (2012).
[5] J. Nocedal, S. Wright: Numerical Optimization. 2nd ed. Springer Series in Op. R. and Fin. Eng. (2006).
[6] A.J. Sommese, J. Verschelde, C.W. Wampler: Introduction to Numerical Algebraic Geometry. Springer (2005).
[7] M. Burr, M. Byrd, K. Lee: Certified algebraic curve projections by path tracking. ISSAC (2025).
[8] N. Boumal. An introduction to optimization on smooth manifolds. Cambridge University Press (2023).
[9] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds [...]. SIAM J. Opt. **22.2** (2012).
[10] J. Milnor: Singular Points of Complex Hypersurfaces. Princeton University Press (1968).