**Riemannian preconditioning**

**Bamdev Mishra** * and **Rodolphe Sepulchre** ††

* Department of Electrical Engineering and Computer Science, University of Liège, 4000 Liège, Belgium (B.Mishra@ulg.ac.be)
† University of Cambridge, Department of Engineering, Cambridge CB2 1PZ, UK (R.Sepulchre@eng.cam.ac.uk)

**Introduction**

- Gradient algorithms are a method of choice for large-scale constrained optimization but their convergence properties critically depend on the metric chosen.
- For optimization problems with equality constraints, sequential quadratic programming (SQP) methods provide an efficient algorithmic procedure based on a local quadratic approximation of the problem. This is the Lagrangian approach.
- An alternative is to embed the constraint into the search space leading to the Riemannian optimization framework that has gained much popularity in the recent years, in particular for orthogonality and rank constraints.

**Contributions**

- We show that SQP provides a systematic framework to choose a metric in Riemannian optimization in a way that takes into consideration both the cost function and the constraints.
- We view this approach of selecting a metric from SQP as a form of Riemannian preconditioning.
- As a first example, the specific situation of quadratic cost and orthogonality constraint is discussed revisiting the classical eigenvalue problem, and connecting to a number of well-known algorithms.
- As the second example, the case of quadratic cost and rank constraint is discussed with applications to matrix Lyapunov equations.

**The constrained optimization viewpoint (SQP)**

- The SQP algorithm (primal form) for \( \min f(x) \) subject to \( h(x) = 0 \)
  1. Compute the search direction \( \xi \) that is the solution to \( \arg \min_{\xi} f(x) + \langle f'(x), \xi \rangle + \frac{1}{2} \langle \xi, D^2 f(x) \xi \rangle \) subject to \( D h(x) \xi = 0 \)
  2. The next iterate \( x \) is obtained by projecting \( x + \xi \) onto the constraint set.
  3. Repeat until convergence.
- Lagrangian \( L(x, \lambda) = f(x) - \langle \lambda, h(x) \rangle \), where \( \lambda \) is the least-square estimate.
- The convergence properties rely on the regularity (non-singularity) of \( L_n(x, \lambda) \) which may not be possible in many applications with underlying symmetries or invariants.
- This is where the quotient manifold framework comes into play.

**Connecting SQP to the Riemannian framework**

- **SQP**: \( \arg \min_{\xi \in \mathbb{R}^n} f(x) + \langle f'(x), \xi \rangle + \frac{1}{2} \langle \xi, D^2 f(x) \xi \rangle \) subject to \( D h(x) \xi = 0 \)
- Riemannian: \( \min_{\xi \in \mathcal{M}} f(x) + \langle f'(x), \xi \rangle + \frac{1}{2} \langle \xi, D^2 f(x) \xi \rangle \)
- Proposition: In the neighborhood of the minimum, the Riemannian gradient descent algorithm is equivalent to SQP algorithm when the metric \( g_0(\xi, \eta) = \langle \xi, D^2 f(x) \eta \rangle \) for all \( \xi, \eta \) in the horizontal space \( \mathcal{H} \).
- Remark: \( L_n(x, \lambda) \) induces a proper metric on the quotient space.

**Quadratic optimization with orthogonality constraint**

\[
\begin{align*}
\min_{X \in O(n)} & \quad \frac{1}{2} \text{Trace}(X'AX) \\
\text{subject to} & \quad X'X = I
\end{align*}
\]

- Optimization is on the Grassmann manifold as the cost remains constant under \( X \mapsto O \times X \)
- Depending on \( A > 0 \) or \( A < 0 \), the cost-related and constraint-related terms are weighted with \( \omega \in [0, 1] \) that is updated dynamically, similar to numerical shifts.
- In all the cases, we propose a family of metrics that generalize power, inverse power, and Rayleigh quotient iterations.
- Each interpreted as a Riemannian steepest descent algorithm with a specific metric.