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*
- For optimization problems with equality constraints, sequential quadratic programming (SQP) methods provide an efficient algorithmic procedure based 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

$$\min_{x\in\mathbb{R}^n} \; f(x) \quad \text{subject to } h(x)=0$$

Sequential quadratic programming (constraints are embedded into the cost function)

$$\min_{x,\lambda} \; f(x) - \langle \lambda, h(x)\rangle,$$

where λ is the Lagrange multiplier

Riemannian framework (constraints are encoded into the seach space)

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

where $\mathcal{M}=\{x : h(x)=0\}$

- A current limitation of Riemannian optimization is in the choice of the metric. Previous work has mostly focused on the search space, exploiting the differential geometry of the constraint but disregarding the role of the cost function

Contributions

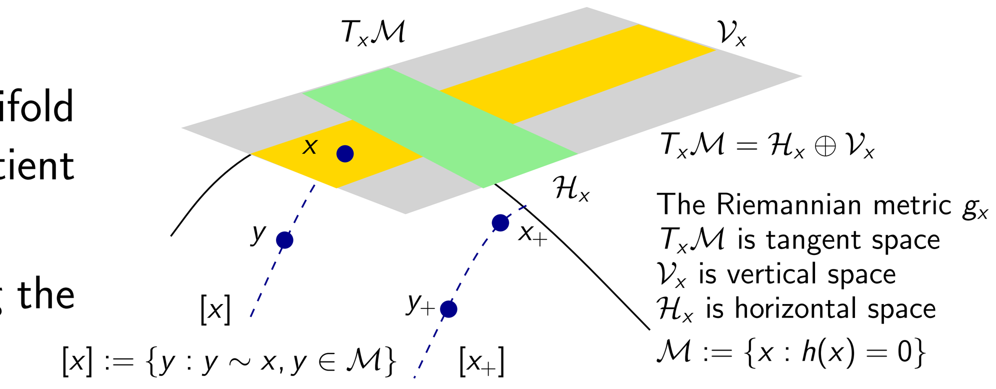
- We show that SQP provides a systematic framework to choose a metric in Riemannian optimization in a way that takes into consideration both the cost function and the constraints
- We view this approach of selecting a metric from SQP as a form of *Riemannian preconditioning*
- As a first example, the 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 $\min_{x\in\mathbb{R}^n} f(x)$ subject to $h(x)=0$
  1. Compute the search direction ζ_x^* that is the solution to
  $$\arg\min_{\zeta_x\in\mathbb{R}^n} \; f(x)+\langle f_x(x),\zeta_x\rangle+\tfrac{1}{2}\langle\zeta_x, \mathrm{D}^2\mathcal{L}(x,\lambda_x)[\zeta_x]\rangle \quad \text{subject to } \mathrm{D}h(x)[\zeta_x]=0$$
  2. The next iterate x_+ is obtained by projecting $x+\zeta_x^*$ onto the constraint set
  3. Repeat until convergence
- Lagrangian $\mathcal{L}(x,\lambda_x)=f(x)-\langle\lambda,h(x)\rangle$, where λ_x is the *least-square estimate*
- The convergence properties rely on the regularity (non-singularity) of $\mathcal{L}_{xx}(x,\lambda)$ which may not be possible in many applications with underlying *symmetries* or *invariances*
- This is where the quotient manifold optimization framework comes into play

The Riemannian optimization viewpoint

- The Riemannian steepest-descent algorithm for $\min_{x\in\mathcal{M}} f(x)$
  1. Search direction: compute the Riemannian gradient $\xi_x=-\mathrm{grad}_x f$ with respect to the Riemannian metric g_x, i.e.,
  $$\arg\min_{\zeta_x\in T_x\mathcal{M}} \; f(x)+\langle f_x(x),\zeta_x\rangle+\tfrac{1}{2}g_x(\zeta_x,\zeta_x)$$
  2. The next iterate is computed using the *retraction*, equivalent to projection
  3. Repeat until convergence
- A well defined scheme on manifold with symmetries, e.g., the quotient manifold
- By construction, $\mathrm{grad}_x f$ is along the horizontal space $\mathcal{H}_x$



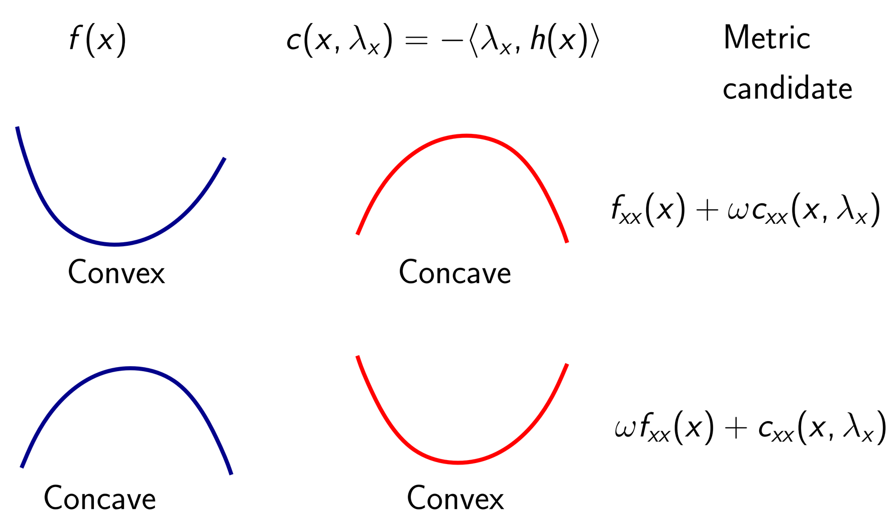
Connecting SQP to the Riemannian framework

- SQP : $\arg\min_{\zeta_x\in\mathbb{R}^n} \; f(x)+\langle f_x(x),\zeta_x\rangle+\tfrac{1}{2}\langle\zeta_x,\mathrm{D}^2\mathcal{L}(x,\lambda_x)[\zeta_x]\rangle$ subject to $\mathrm{D}h(x)[\zeta_x]=0$

  Riemannian : $\arg\min_{\zeta_x\in T_x\mathcal{M}} \; f(x)+\langle f_x(x),\zeta_x\rangle+\tfrac{1}{2}g_x(\zeta_x,\zeta_x)$
- Proposition: In the neighborhood of the minimum, the Riemannian gradient descent algorithm is equivalent to SQP algorithm when the metric $g_x(\zeta_x,\eta_x)=\langle\zeta_x,\mathrm{D}^2\mathcal{L}(x,\lambda_x)[\eta_x]\rangle$ for all ζ_x,η_x in the horizontal space $\mathcal{H}_x$
- Remark: $\mathcal{L}_{xx}(x,\lambda_x)$ induces a proper metric on the quotient space

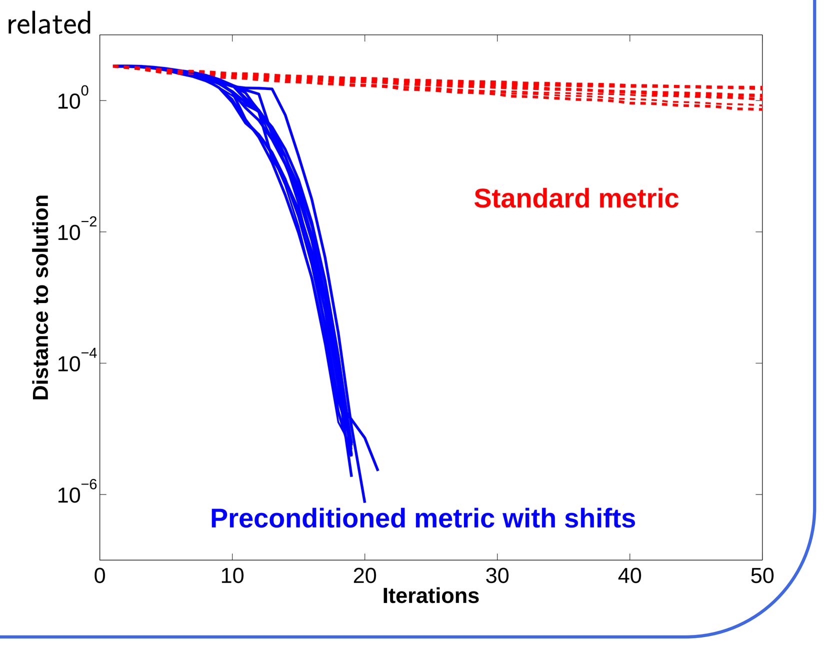
Riemannian optimization and local convexity

- This problem is well-defined
  $$\arg\min_{\zeta_x\in\mathcal{H}_x} f(x)+\langle f_x(x),\zeta_x\rangle + \tfrac{1}{2}\langle\zeta_x,\mathrm{D}^2\mathcal{L}(x,\lambda_x)[\zeta_x]\rangle$$
- $$g_x(\xi_x,\eta_x)=\langle\xi_x,\mathrm{D}^2\mathcal{L}(x,\lambda_x)[\eta_x]\rangle = \underbrace{\langle\xi_x,\mathrm{D}^2 f(x)[\eta_x]\rangle}_{\text{cost related}} + \underbrace{\langle\xi_x,\mathrm{D}^2 c(x,\lambda_x)[\eta_x]\rangle}_{\text{constraint related}},$$



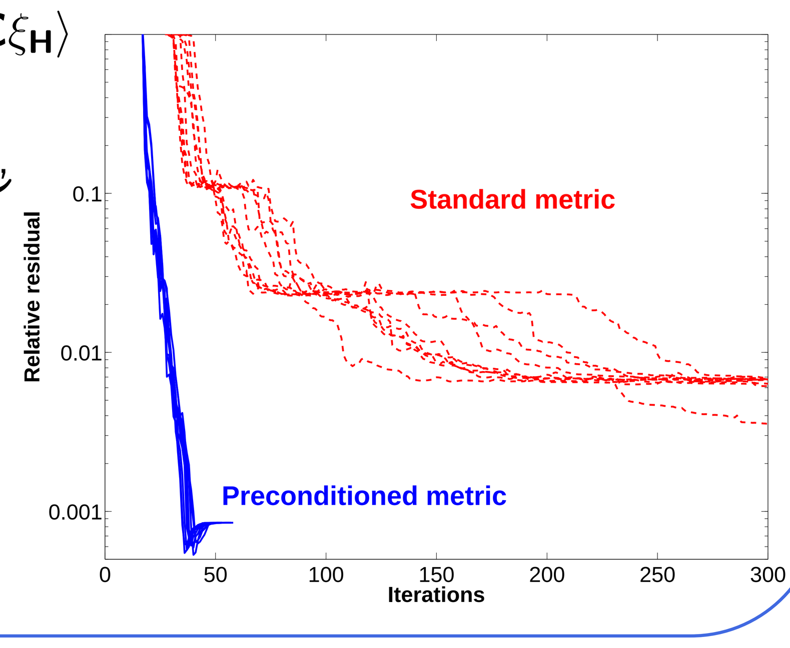
Quadratic optimization with orthogonality constraint

- $\min_{\mathbf{X}\in\mathbb{R}^{n\times r}} \; \frac{1}{2}\mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X})$ subject to $\mathbf{X}^T\mathbf{X}=\mathbf{I}$
- Optimization is on the Grassmann manifold as the cost remains constant under $\mathbf{X}\mapsto\mathbf{XO}$, $\mathbf{O}$ is a $r\times r$ orthogonal matrix
- $$\begin{aligned}\mathcal{L}(x,\lambda_x) &= \mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X})/2-\langle\lambda_x,\mathbf{X}^T\mathbf{X}-\mathbf{I}\rangle \text{ with } \lambda_x=\mathbf{X}^T\mathbf{A}\mathbf{X}\\ \Rightarrow \mathrm{D}^2\mathcal{L}(x,\lambda_x)[\xi_x] &= \mathbf{A}\xi_x-\xi_x\lambda_x,\\ g_x(\xi_x,\eta_x) &= \underbrace{\langle\xi_x,\mathbf{A}\eta_x\rangle}_{\text{cost related}}-\underbrace{\langle\xi_x,\eta_x\mathbf{X}^T\mathbf{A}\mathbf{X}\rangle}_{\text{constraints related}}\end{aligned}$$
- Depending on $\mathbf{A}\succ 0$ or $\mathbf{A}\prec 0$, the cost-related and constraint-related terms are weighed 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



Quadratic optimization with rank constraint

- $\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \; \frac{1}{2}\mathrm{Trace}(\mathbf{X}^T\mathbf{AXB})+\mathrm{Trace}(\mathbf{X}^T\mathbf{C})$ subject to $\mathrm{rank}(\mathbf{X})=r$,
- Fixed-rank parameterization $\mathbf{X}=\mathbf{GH}^T$, where $\mathbf{G}\in\mathbb{R}_*^{n\times r}$ (full column rank matrices) and $\mathbf{H}\in\mathbb{R}_*^{m\times r}$ and cost is constant under $(\mathbf{G},\mathbf{H})\mapsto(\mathbf{GM}^{-1},\mathbf{HM}^T)$, $\mathbf{M}$ is $r\times r$ non-singular matrix
- Exploiting the fact that the cost is quadratic in arguments $\mathbf{G},\mathbf{H}$ individually, we propose a family of Riemannian metrics parameterized by $\omega\in[0,1)$
- $$g_x(\xi_x,\eta_x)=\omega\langle\eta_{\mathbf{G}},2\mathbf{AG}\mathrm{Sym}(\mathbf{H}^T\mathbf{B}\xi_{\mathbf{H}})+\mathbf{C}\xi_{\mathbf{H}}\rangle +\omega\langle\eta_{\mathbf{H}},2\mathbf{BH}\mathrm{Sym}(\mathbf{G}^T\mathbf{A}\xi_{\mathbf{G}})+\mathbf{C}^T\xi_{\mathbf{G}}\rangle + \underbrace{\langle\eta_{\mathbf{G}},\mathbf{A}\xi_{\mathbf{G}}\mathbf{H}^T\mathbf{BH}\rangle+\langle\eta_{\mathbf{H}},\mathbf{B}\xi_{\mathbf{H}}\mathbf{G}^T\mathbf{AG}\rangle}_{\text{Block diagonal approximation of } f_{xx}(x)},$$
- $\omega=0$ is a good choice for the metric as $\mathbf{H}^T\mathbf{BH}$ and $\mathbf{G}^T\mathbf{AG}$ are $\succ 0$ for all $(\mathbf{G},\mathbf{H})\in\mathbb{R}_*^{n\times r}\times\mathbb{R}_*^{m\times r}$
- The family of metrics connect to metrics in state-of-the-algorithms for low-rank matrix completion and Lyapunov equations



This poster presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. Bamdev Mishra is supported as an F.R.S.-FNRS research fellow (Belgian Fund for Scientific Research).