Riemannian adaptive stochastic gradient algorithms on matrix manifolds

Hiroyuki Kasai†, Pratik Jawanpuria*, and Bamdev Mishra*

† The University of Electro-Communications, Japan
* Microsoft, India

Pacific Ballroom #108
Optimization on manifolds

\[ \min_{x \in \mathcal{M}} f(x) \]

\(\mathcal{M}\) is smooth manifold.
\(f\) is smooth function.

Manifolds of interest:
Grassmann / subspaces
Low-rank matrices / tensors
Hyperbolic spaces
Positive definite matrices
Stochastic update on manifold:

\[ x_{t+1} = R_{x_t}(-\alpha_t \nabla f_t(x_t)) \]
Adaptive algorithms on manifolds

Euclidean adaptive update:

$$x_{t+1} = x_t - \alpha_t \underbrace{V_t^{-1/2}}_{\text{adaptive scaling}} \nabla f_t(x_t),$$

Riemannian adaptive update:

$$x_{t+1} = R_{x_t}(-\alpha_t \underbrace{V_t^{-1/2}}_{\text{adaptive scaling}} \text{grad} f_t(x_t)), $$
The challenge is how to compute $V$ in the Riemannian setting.
Our contributions:

• Propose a principled approach for modeling adaptive weights for Riemannian stochastic gradient, i.e., model adaptive weight matrices for row and column subspaces exploiting the geometry of the manifold.

• Provide convergence analysis of our algorithm, under a set of mild conditions. Our algorithms achieve a rate of convergence order $O(\log(T)/\sqrt{T})$, where $T$ is the number of iterations, for non-convex stochastic optimization.
Riemannian adaptive stochastic gradient algorithms on matrix manifolds

Hiroyuki Kasai†, Pratik Jawanpuria*, and Bamdev Mishra*

† The University of Electro-Communications, Japan
* Microsoft, India

Pacific Ballroom #108