The extension of DRSOM on manifold

Speaker: Tianyun Tang

August 4, 2022

Codes were written with the help of Dr Nachuan Xiao.

From unconstrained optimization to manifold optimization

Unconstrained optimization

$$\min \{f(x): \; x \in \mathbb{R}^n\}. \tag{1}$$

Algorithms: Let $g_k := \nabla f(x_k)$, $H_k := \nabla^2 f(x_k)$ and $d_k := x_k - x_{k-1}$.

1, Gradient descent: $x_{k+1} := x_k - \alpha_k \nabla g_k$.

2, Newton method: $x_{k+1} := x_k - H_k^{-1} g_k$

3, DRSOM(Prof. Ye et al. 2022): $x_{k+1} := x_k - \alpha_k^1 g_k + \alpha_k^2 d_k$, and

$$\alpha_k := \arg \min_{\|\alpha_k\|_{G_k} \leq \triangle} m_k^\alpha(\alpha),$$

where the 2-dimensional quadratic model $m_k^\alpha(\alpha)$ is defined as
$m_k^\alpha(\alpha) := f(x_k) + c_k^\top \alpha + \frac{1}{2}\alpha^\top Q_k \alpha$

$$Q_k = \begin{bmatrix} g_k^\top H_k g_k & -d_k^\top H_k g_k \\ -d_k^\top H_k g_k & d_k^\top H_k d_k \end{bmatrix}, \; c_k = \begin{bmatrix} -\|g_k\|^2 \\ g_k^\top d_k \end{bmatrix}, \; G_k := \begin{bmatrix} g_k^\top g_k & -d_k^\top g_k \\ -d_k^\top g_k & d_k^\top d_k \end{bmatrix}.$$

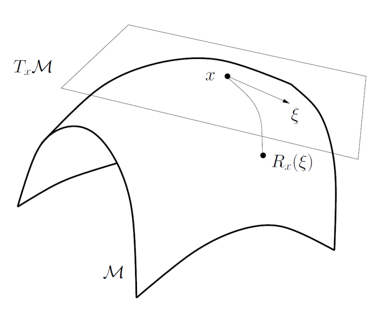
Question: Can we extend DRSOM to manifold optimization?

Manifold optimization

$$\min \{f(x): \; c(x) = 0, \; x \in \mathbb{R}^n\}. \tag{2}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is smooth.
- $c : \mathbb{R}^n \to \mathbb{R}^p$ is smooth and $J_c(x) := [\nabla c_1(x), \ldots, \nabla c_p(x)] \in \mathbb{R}^{n \times p}$ has full rank for any $x \in \mathcal{M} := \{x \in \mathbb{R}^n : \; c(x) = 0\}$.

$\Rightarrow$ $\mathcal{M}$ is a Riemannian manifold in $\mathbb{R}^n$.



Manifold optimization

$$\text{Tangent space}: \;\; T_x\mathcal{M} := \left\{ d \in \mathbb{R}^n : \;\; d^\top J_c(x) = 0 \right\}. \tag{3}$$

$$\text{Projection}: \;\; \mathrm{Proj}_x[d] := d - J_c(x) J_c(x)^\dagger d. \tag{4}$$

$$\text{Closest point retraction}: \;\; \mathcal{R}_\mathcal{M}(x) \in \arg\min \left\{ \|y - x\| : \;\; y \in \mathcal{M} \right\}. \tag{5}$$

$$\text{Riemannian gradient}: \;\; \mathrm{grad} f(x) := \mathrm{Proj}_x[\nabla f(x)]. \tag{6}$$

$$\text{Riemannian hessian}: \;\; \mathrm{Hess} f(x) : h \to \mathrm{Proj}_x \left[\mathrm{D} \left(\widetilde{\mathrm{grad} f(x)} \right) [h] \right]. \tag{7}$$

Manifold optimization

Algorithms:

Let $g_k := \mathrm{grad} f(x_k)$, $H_k := \mathrm{Hess} f(x_k)$ and $d_k := \mathrm{Proj}_{x_k}(x_k - x_{k-1})$.

1, Riemannian gradient descent: $x_{k+1} := \mathcal{R}_{\mathcal{M}}(x_k - \alpha_k g_k)$.

2, Riemannian newton method: $x_{k+1} := \mathcal{R}_{\mathcal{M}}\left(x_k - H_k^{-1} g_k\right)$.

3, Riemannian DRSOM: $x_{k+1} := \mathcal{R}_{\mathcal{M}}\left(x_k - \alpha_k^1 g_k + \alpha_k^2 d_k\right)$, and

$$\alpha_k := \arg \min_{\|\alpha_k\|_{G_k} \le \triangle} m_k^\alpha(\alpha),$$

where the 2-dimensional quadratic model $m_k^\alpha(\alpha)$ is defined as
$m_k^\alpha(\alpha) := f(x_k) + c_k^\top \alpha + \frac{1}{2}\alpha^\top Q_k \alpha$

$$Q_k = \begin{bmatrix} g_k^\top H_k g_k & -d_k^\top H_k g_k \\ -d_k^\top H_k g_k & d_k^\top H_k d_k \end{bmatrix}, \; c_k = \begin{bmatrix} -\|g_k\|^2 \\ g_k^\top d_k \end{bmatrix}, \; G_k := \begin{bmatrix} g_k^\top g_k & -d_k^\top g_k \\ -d_k^\top g_k & d_k^\top d_k \end{bmatrix}.$$

Numerical experiments

Max-cut SDP

$$\text{Max cut}: \ \min\left\{\left\langle C, RR^\top\right\rangle : \ \mathrm{diag}(RR^\top) = e, \ R \in \mathbb{R}^{n\times r}\right\}. \tag{8}$$

We terminate algorithms when $\|\mathrm{grad} f(R)\| < 10^{-6}$.

problem	algorithm	Rp	Rd	pdgap	obj	time
g1	DRSOM	2.30e-16	5.95e-10	1.51e-16	-4.8332791e+04	**1.20e+00**
n=800	TR	2.75e-16	4.14e-13	1.51e-16	-4.8332791e+04	1.95e+00
m=19176	CG	2.59e-16	1.84e-10	1.51e-16	-4.8332791e+04	3.32e+00
	BFGS	2.46e-16	6.85e-10	2.26e-16	-4.8332791e+04	6.72e+00
g32	DRSOM	3.08e-16	2.92e-10	2.18e-16	-6.2705586e+03	**9.29e+00**
n=2000	TR	2.47e-16	1.25e-09	0.00e+00	-6.2705586e+03	4.42e+01
m=4000	CG	2.54e-16	9.01e-10	7.25e-17	-6.2705586e+03	2.33e+01
	BFGS	2.55e-16	4.30e-10	7.25e-17	-6.2705586e+03	7.57e+01
g48	DRSOM	3.33e-16	3.03e-15	2.12e-15	-2.4000000e+04	2.02e+00
n=3000	TR	2.67e-16	1.67e-16	7.58e-17	-2.4000000e+04	**1.96e+00**
m=6000	CG	2.85e-16	2.80e-14	5.31e-16	-2.4000000e+04	2.80e+00
	BFGS	2.84e-16	1.02e-14	2.27e-16	-2.4000000e+04	8.71e+00
g55	DRSOM	3.74e-16	4.83e-10	8.24e-17	-4.4157842e+04	**8.91e+00**
n=5000	TR	2.99e-16	8.37e-11	8.24e-17	-4.4157842e+04	9.89e+00
m=12498	CG	3.03e-16	1.05e-10	4.12e-16	-4.4157842e+04	1.51e+01
	BFGS	3.02e-16	6.86e-10	4.94e-16	-4.4157842e+04	4.16e+01
g67	DRSOM	4.31e-16	9.65e-11	4.70e-16	-3.0977746e+04	**1.21e+02**
n=10000	TR	3.36e-16	1.68e-10	9.98e-16	-3.0977746e+04	1.03e+03
m=20000	CG	3.43e-16	4.02e-10	4.70e-16	-3.0977746e+04	2.13e+02
	BFGS	3.41e-16	1.48e-10	2.35e-16	-3.0977746e+04	1.04e+03

Discretized 1D Kohn-Sham Equation

$$\min\left\{\frac{1}{2}\operatorname{tr}\left(R^\top L R\right)+\frac{\alpha}{4}\operatorname{diag}(RR^\top)^\top L^{-1}\operatorname{diag}(RR^\top):\ R^\top R=I_p,\ R\in\mathbb{R}^{n\times r}\right\}, \tag{9}$$

where L is a tri-diagonal matrix with 2 on its diagonal and -1 on its subdiagonal and $\alpha > 0$ is a parameter. We terminate algorithms when $\|\mathrm{grad} f(R)\| < 10^{-4}$.

problem	algorithm	gradnorm	obj	time
$n = 1000$	DRSOM	8.96e-05	2.1070857e+02	**1.68e-01**
$r = 20$	TR	7.49e-05	2.1070857e+02	9.99e-01
$\alpha = 1$	CG	9.09e-05	2.1070857e+02	6.82e-01
	BFGS	9.95e-05	2.1070857e+02	7.01e+00
$n = 1000$	DRSOM	9.67e-05	2.8107086e+03	**5.95e-01**
$r = 50$	TR	1.21e-05	2.8107086e+03	2.23e+00
$\alpha = 1$	CG	9.95e-05	2.8107086e+03	2.14e+00
	BFGS	9.63e-05	2.8107086e+03	2.00e+01
$n = 10000$	DRSOM	9.56e-05	2.1070857e+02	**5.68e-01**
$r = 20$	TR	1.04e-05	2.1070857e+02	1.81e+00
$\alpha = 1$	CG	8.46e-05	2.1070857e+02	1.71e+00
	BFGS	9.20e-05	2.1070857e+02	1.10e+01
$n = 10000$	DRSOM	9.91e-05	2.8107086e+03	**4.39e+00**
$r = 50$	TR	3.64e-05	2.8107086e+03	6.59e+00
$\alpha = 1$	CG	9.46e-05	2.8107086e+03	8.43e+00
	BFGS	9.43e-05	2.8107086e+03	4.37e+01