Convex Optimization Methods for Dimension Reduction and Coefficient Estimation in Multivariate Linear Regression

Renato Monteiro (Georgia Tech)

FODAVA KICKOFF MEETING

Atlanta, USA

Sept. 16, 2008

P.I.'s Research Outline and Goals

P.I.'s research lies in the area of continuous optimization:

- **cone programming including linear, second-order and semidefinite programming**
- **numerical algorithms including interior-point and first-order methods for large scale optimization problems**
- **applications to statistics and graph theory**

Research Goals:

- **develop efficient algorithms for solving optimization problems that arise in the context of data analytics and visualization**
- **work on the mathematical foundation of sparse data representation and recovery (e.g., compressed sensing, dimension reduction)**

CONE PROGRAMMING

Given n-dimensional vector space $\Re^n$ and a closed convex cone $\mathcal{K} \subseteq \Re^n$, the CP problem is:

$$\min\{\langle c, x\rangle : \mathcal{A}(x) = b,\ x \in \mathcal{K}\}$$

where $c \in \Re^n$, $b \in \Re^m$ and $\mathcal{A} : \Re^n \to \Re^m$ is a linear map. Its dual is

$$\max\{\langle b, y\rangle : c - \mathcal{A}^*(y) \in \mathcal{K}^*\}$$

where $\mathcal{K}^* := \{s \in \Re^n : \langle s, x\rangle \geq 0,\ \forall x \in \mathcal{K}\}$.

Algorithms:

- interior-point (IP) methods (for most important CP's)
- first-order methods for large instances

IP methods: advantage: few iterations; highly accurate solutions; good codes available — disadvantage: high iteration cost and memory requirement

first-order methods: disadvantage: many iterations; low to medium accurate solutions; no general purpose code available — advantage: low iteration cost and memory requirement

Min-max or saddle point problems

Their general form is

$$\min_{x\in X}\left(\Phi(x) := \max_{y\in Y}\phi(x,y)\right)$$

where X, Y are simple closed convex sets, ϕ is convex in x and concave in y.

Under the assumption that $\nabla\phi$ is Lipschitz continuous, first-order methods with known iteration-complexity bounds have been developed to solve these problems:

- **Nesterov's smooth or nonsmooth methods and their variants;**
- **Nemirovski's prox-mirror method**

Dimension Reduction in Statistics

Assume that $B \in \Re^{n\times q}$ collects n observations on q responses $\mathbf{b} = (b_1, \ldots, b_q)'$, and $A \in \Re^{n\times p}$ consists of n observations on p explanatory variables $\mathbf{a} = (a_1, \ldots, a_p)'$. Assume also that $q < p$. Consider the multivariate linear model

$$B = AU + E,$$

where $U \in \Re^{p\times q}$ is a coefficient matrix, $E = (\mathbf{e^1}, \ldots, \mathbf{e^n})'$ is the regression noise, and all $\mathbf{e^i}$s are indep sampled from $\mathcal{N}(0, \Sigma)$.

To estimate U and accomplish dimension reduction, Yuan et al. proposed to solve

$$\min_U \frac{1}{2}\|AU - B\|_F^2 + \lambda \sum_{i=1}^{q} \sigma_i(U) \qquad (1)$$

for different $\lambda > 0$ values. Here, $\|V\|_F^2 := \sum_{i,j} V_{ij}^2$ and $\sigma_i(U) :=$ i-th largest singular value of U.

Optimal solutions U of (1) tend to have low rank ($<< \min(p, q)$) and hence sparse SVD.

A paper by Lu, M. and Yuan presents a couple of reformulations of (1) as either a CP or min-max problem.

It also discusses the performance of a proper (either IP or first-order) algorithm for solving each reformulation.

Cone programming reformulation

Problem (1) can be formulated as a cone programming.

Write $V \succeq 0$ if V is symmetric and positive semidefinite. Also, let $\mathcal{S}^l$ denote the space of $l \times l$ symm. matrices.

Proposition: Let $U \in \Re^{p\times q}$, $k \le \min\{p, q\}$ be given and set $l := p + q$. For $t \in \Re$, we have

$$\sum_{i=1}^{k} \sigma_i(U) \le t \;\Leftrightarrow\; \left\{ \begin{array}{rcl} t - ks - \mathrm{Trace}(V) & \geq & 0, \\ V - \mathcal{G}(U) + sI & \succeq & 0, \\ V & \succeq & 0, \end{array} \right.$$

for some $V \in \mathcal{S}^l$ and $s \in \Re$, where $\mathcal{G} : \Re^{p\times q} \to \Re^{l\times l}$ is defined as

$$\mathcal{G}(U) := \begin{pmatrix} 0 & U^T \\ U & 0 \end{pmatrix}$$

MIN-MAX REFORMULATIONS

Problem (1) can also be reformulated as

$$\max_{W \in \Omega} \min_{\|U\|_F \leq r} \left\{ \frac{1}{2}\|AU - B\|_F^2 + \lambda q \, \mathcal{G}(U) \bullet W \right\}, \quad (2)$$

where r **is an appropriate scalar and**

$$\Omega \quad := \quad \{W \in \mathcal{S}^{p+q} : 0 \preceq W \preceq I/q, \, \mathrm{Trace}(W) = 1\}$$

Proposition: Assume A **has full column rank. Given** $\epsilon > 0$**, Nesterov's smooth method finds an** ϵ**-optimal solution of (2) in a number of iterations which does not exceed**

$$\frac{2\lambda\|(A^T A)^{-1/2}\|}{\sqrt{\epsilon}} \sqrt{q \log\left(\frac{p+q}{q}\right)}.$$

Note: The complexity of solving the min-max version of (2) is $\mathcal{O}(1/\epsilon)$ **instead of** $\mathcal{O}(1/\sqrt{\epsilon})$ **as above.**

Computational Results

The entries of $A \in \Re^{n \times p}$ and $B \in \Re^{n \times q}$, with $p = 2q$ and $n = 10q$, were uniformly generated in [0,1]. The accuracy in the table below is $\epsilon = 10^{-1}$.

Problem	Iter		Time	
(p, q)	MIN-MAX	MAX-MIN	MIN-MAX	MAX-MIN
(200, 100)	610	1	29.60	0.91
(400, 200)	1310	1	432.92	8.36
(600, 300)	2061	1	2155.76	31.23
(800, 400)	2848	1	7831.09	76.75
(1000, 500)	3628	1	21128.70	156.68
(1200, 600)	4436	1	47356.32	276.64
(1400, 700)	5280	1	98573.73	456.61
(1600, 800)	6108	1	176557.49	699.47

The table below compares the max-min formulation (M-MIN) with the cone programming reformulation. The accuracy is $\epsilon = 10^{-8}$.

Problem	Iter		Time		Memory	
(p, q)	M-MIN	CONE	M-MIN	CONE	M-MIN	CONE
(20,10)	3455	17	3.61	5.86	2.67	279
(40,20)	1696	15	6.90	77.25	2.93	483
(60,30)	1279	15	13.33	506.14	3.23	1338
(80,40)	1183	15	25.34	2205.13	3.63	4456
(100,50)	1073	19	40.66	8907.12	4.23	10445
(120,60)	1017	N/A	62.90	N/A	4.98	> 16109

Summary:

I plan to bring to the attention of the data and visual analytics community existing optimization tools that can speed up the solution of the optimization models used in their work.

I also plan to develop new optimization tools to better exploit special structures of some classes of optimization models arising in the course of our research in the area of data and visual analytics.

It is also my intent to contribute to other topics of data and visual analytics as I learn more about them.