Semidefinite Programming	**Lecture 28**
OR 6327 Spring 2012	**May 3, 2012**
Scribe: Joyjit Roy	

Matrix Completion Problem (aka the Netflix Problem)

Given **some** entries of an $m \times n$ matrix $\mathbf{M}$, say those for $ij \in \Omega$, we wnat to recover $\mathbf{M}$. If $|\Omega| < mn$, the problem is obviously underdetermined. However, if we know that $\mathbf{M}$ has low rank, maybe we can recover it. This can be formulated as the problem:

$$\begin{array}{rrcl} \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} & \operatorname{rank}(\mathbf{X}) & & \\ & x_{ij} & = & m_{ij}, \quad ij \in \Omega. \end{array}$$

Consider the convex relaxation of this problem:

$$\begin{array}{rrcl} \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} & \|\mathbf{X}\|_* & & \\ & x_{ij} & = & m_{ij}, \quad ij \in \Omega. \end{array}$$

Recall that $\|\mathbf{X}\|_* = \|\sigma(\mathbf{X})\|_1$. This problem can be modelled as an SDP but it is hard to solve using interior-point methods so we look to solve this directly.
Cai, Candes & Shen proposed a first-order method to solve the related problem:

$$\begin{array}{rrcl} \min_{\mathbf{X}\in\mathbb{R}^{m\times n}} & \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X}\|_F^2 & & \\ & x_{ij} & = & m_{ij}, \quad ij \in \Omega. \end{array}$$

With the intention of using a Lagrangian method we introduce the function

$$L(\mathbf{X},\mathbf{Y}) := \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X}\|_F^2 - \sum_{ij\in\Omega} y_{ij}(x_{ij} - m_{ij}).$$

Here $\mathbf{Y} \in \mathbb{R}^{m\times n}$, but only its ij entries for $ij \in \Omega$ are relevant.
For any $\mathbf{Z} \in \mathbb{R}^{m\times n}$, let $\mathbf{Z}_\Omega$ denote the matrix with entries z_{ij}, if $ij \in \Omega$ and 0, if $ij \notin \Omega$. Then

$$\begin{aligned} L(\mathbf{X},\mathbf{Y}) &= \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X}\|_F^2 - \mathbf{Y}_\Omega \bullet (\mathbf{X} - \mathbf{M}) \\ &= \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}_\Omega\|_F^2 + \mathbf{Y}_\Omega \bullet \mathbf{M} - \frac{1}{2}\|\mathbf{Y}_\Omega\|_F^2 \\ &= \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X}\|_F^2 - \mathbf{Y} \bullet (\mathbf{X} - \mathbf{M})_\Omega. \end{aligned}$$

This leads to the following algorithm:

Algorithm Choose $\mathbf{X}_0, \mathbf{Y}_0$ and $\{\delta_k\}_{k=0}^{\infty}$.
Iteration k $\mathbf{X}_{k+1} := \arg\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} L(\mathbf{X}, \mathbf{Y}_k)$,
$\mathbf{Y}_{k+1} := \mathbf{Y}_k - \delta_k(\mathbf{X}_{k+1} - \mathbf{M})_\Omega$.

This is in fact Uzawa's method going back to the 1950's. Convergence of the algorithm can be proved if $0 < \delta_k < 2$, all k.
We now look at the question of solving the subproblems efficiently.

Lemma 1 *If* $\mathbf{Z} \in \mathbb{R}^{m \times n}$*, then* $\|\mathbf{Z}\|_* \geq \|\mathrm{diag}\,(\mathbf{Z})\|_1$ *with equality if* $\mathbf{Z}$ *is diagonal.*

Proof: If $\mathbf{Z}$ is diagonal, then $\sigma(\mathbf{Z}) = |\mathrm{diag}\,(\mathbf{Z})|$. So equality holds trivially.
In general, suppose $\|\mathbf{Z}\|_* = 1$.
So,

$$\mathbf{Z} = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^T = \sum_{i=1}^{r} \sigma_i \mathbf{p}_i \mathbf{q}_i^T,$$

where $\mathbf{P} \in \mathbb{R}^{m \times m}$ and $\mathbf{Q} \in \mathbb{R}^{n \times n}$ are orthogonal, with columns $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m$ and $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ respectively, $\Sigma =$ "Diag $(\sigma_1, \sigma_2, \ldots, \sigma_r)$" $\in \mathbb{R}^{m \times n}$ and $r := \min\{m, n\}$.
So $\mathbf{Z}$ is a convex combination of rank 1 matrices of the form $\mathbf{p}\mathbf{q}^T$. Observe that

$$\begin{aligned} \|\mathrm{diag}\,(\mathbf{p}\mathbf{q}^T)\|_1 &= |\mathbf{p}_1||\mathbf{q}_1| + |\mathbf{p}_2||\mathbf{q}_2| + \ldots + |\mathbf{p}_r||\mathbf{q}_r|. \\ &= \text{the inner product of truncations to length } r \text{ of vectors of absolute values of entries of} \end{aligned}$$

By the Cauchy-Schwarz inequality, this is at most the product of their norms, so at most 1.
Hence by the convexity of $\|\cdot\|_1$, $\|\mathrm{diag}\,(\mathbf{Z})\|_1 \leq 1$.

Proposition 1 *If* $\mathbf{Y}_\Omega = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^T$ *is the SVD of* $\mathbf{Y}_\Omega$*, then the unique solution to*

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}_\Omega\|_F^2$$

is given by $\mathbf{X} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}^T$*, where* $\mathbf{\Lambda} = \mathrm{Diag}\,(\lambda)$*,* $\mathbf{\Sigma} = \mathrm{Diag}\,(\sigma)$*, and* $\lambda_j := (\sigma_j - \tau)_+$*, for all* j*.*

Proof: Let $\mathbf{Z} = \mathbf{P}^T\mathbf{X}\mathbf{Q}$, so that

$$\begin{aligned} & \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}_\Omega\|_F^2 \\ = \; & \tau\|\mathbf{Z}\|_* + \frac{1}{2}\|\mathbf{Z} - \mathbf{\Sigma}\|_F^2 \\ \geq \; & \tau\|\mathrm{diag}\,(\mathbf{Z})\|_1 + \frac{1}{2}\|\mathrm{diag}\,(\mathbf{Z}) - \sigma\|_2^2, \end{aligned}$$

with equality holding if $\mathbf{Z}$ is diagonal. So the optimal solution is $\mathbf{Z} = \mathrm{Diag}\,(\lambda)$ where λ minimizes

$$\sum_{j=1}^{r} \tau|\lambda_j| + \frac{1}{2}(\lambda_j - \sigma_j)^2.$$

This is minimized by $\lambda_j = (\sigma_j - \tau)_+$, for all j, as in Proposition 2 of last time.

Remark 1 *If we suspect that* $\mathbf{X}$ *has low rank, we only need the* σ_j*'s ,* p_j*'s and* q_j*'s corresponding to the few* j*'s with* $\sigma_j > \tau$*.*

We now detail some experimental results for matrix completion from Cai, Candes & Shen's paper (Table 5.1). Here the rank r is the rank of the unknown matrix $\mathbf{X}$, m/d_r is the ratio between the number of sampled entries and the number of degrees of freedom in an $n \times n$ matrix of rank r (oversampling ratio), and m/n^2 is the fraction of observed entries. All the computational results are averaged over five runs.

Unknown $\mathbf{X}$				Computational results		
size($m \times n$)	rank(r)	m/d_r	m/n^2	time(s)	# iters	relative error
1000 × 1000	10	6	0.12	23	117	1.64×10^{-4}
	50	4	0.39	196	114	1.59×10^{-4}
	1000	3	0.57	501	129	1.68×10^{-4}
5000 × 5000	10	6	0.024	147	123	1.73×10^{-4}
	50	5	0.10	950	108	1.61×10^{-4}
	100	4	0.158	3339	123	1.72×10^{-4}
10000 × 10000	10	6	0.012	281	123	1.73×10^{-4}
	50	5	0.050	2096	110	1.65×10^{-4}
	100	4	0.080	7059	127	1.79×10^{-4}
20000 × 20000	10	6	0.006	588	124	1.73×10^{-4}
	50	5	0.025	4581	111	1.66×10^{-4}
30000 × 30000	10	6	0.004	1030	125	1.73×10^{-4}

We end with an outline of the course.

- **Applications:**
  - Eigenvalue and SVD problems
  - Control Theory
  - Structural Optimization
  - Relaxations of Max-cut, Lovász theta function
  - Global polynomial optimization
  - Robust optimization
  - High-dimensional statistics and machine learning
- **Theory:**
  - Beautiful duality theory (needs Slater conditions)
    - bounds on optimality gaps
    - suggests good algorithms
- **Algorithms:**

- Interior-point methods
  - theoretically attractive
  - reasonable for reasonably-sized problems ($m \leq 1000, n \leq 1000$)
  - inexact versions for larger problems
- First-order methods
  - Spectral bundle method
  - Low-rank methods
  - BMZ linear transform
- Specialized Algorithms
  - Sparse covariance selection
  - Low-rank matrix completion

This covers *the full spectrum* of optimization.