Advanced Topics in Machine Learning (600.692)
Homework 5: Robust PCA

Instructor: René Vidal
Due Date: 03/28/2014, 11.59PM Eastern

READING MATERIAL: Chapter 3 and Appendix A of GPCA book.

1. (Properties of the $\ell_{2,1}$ Norm).
   
   (a) Let $x$ be a vector. Show that the sub-gradient of the $\ell_2$ norm is given by
   
   \[ \partial \|x\|_2 = \begin{cases} \frac{x}{\|x\|_2} & \text{if } x \neq 0 \\ \{w : \|w\|_2 \leq 1\} & \text{if } x = 0 \end{cases} \]  
   
   (1)

   (b) Let $X$ be a matrix. Show that the $\ell_{2,1}$ norm of $X$, $f(X) = \|X\|_{2,1} = \sum_j \|X_{.,j}\|_2 = \sum_j \sqrt{\sum_i X_{ij}^2}$, is a convex function of $X$.

   (c) Show that the sub-gradient of the $\ell_{2,1}$ norm is given by
   
   \[ (\partial \|X\|_{2,1})_{ij} = \begin{cases} \frac{X_{ij}}{\|X_{.,j}\|_2} & X_{.,j} \neq 0 \\ W_{ij} : \|W_{.,j}\|_2 \leq 1 & X_{.,j} = 0 \end{cases} \]  
   
   (2)

   (d) Show that the optimal solution of
   
   \[ \min_A \frac{1}{2} \|X - A\|_F^2 + \lambda \|A\|_{2,1} \]  
   
   (3)

   is given by $A = XS_\tau(\text{diag}(x))\text{diag}(x)^{-1}$, where $x$ be a vector whose $j$-th entry is given by $x_j = \|X_{.,j}\|_2$, and $\text{diag}(x)$ is a diagonal matrix with the entries of $x$ in its diagonal. By convention, if $x_j = 0$, then the $j$-th entry of $\text{diag}(x)^{-1}$ is also zero.

2. Let $X = L_0 + E_0$ be a matrix formed as the sum of a low rank matrix $L_0$ and a matrix of corruptions $E_0$, where the corruptions can be either outlying entries (gross errors) or outlying data points (outliers).

   (a) (PCA with robustness to outliers). Assuming that the matrix $X$ is fully observed and that the matrix $E_0$ is a matrix of outliers, propose an algorithm for solving the following optimization problem
   
   \[ \min_{L,E} \|L\|_* + \lambda \|E\|_{2,1} \text{ s.t. } X = L + E. \]  
   
   (4)

   (b) (PCA with robustness to missing entries and gross errors). Assuming that you observe only a fraction of the entries of $X$ as indicated by a set $\Omega$ and that the matrix $E_0$ is a matrix of gross errors, propose an algorithm for solving the following optimization problem
   
   \[ \min_{L,E} \|L\|_* + \lambda \|E\|_1 \text{ s.t. } P_\Omega(X) = P_\Omega(L + E). \]  
   
   (5)
3. **Implementation of Reweighted Least Squares and Robust PCA.** Implement the functions below using as few lines of MATLAB code as possible. Compare the performance of these methods by modifying the sample code `test_matrixcompletion.m`. Which method works better and which regime (e.g., depending on percentage of corrupted entries (or corrupted data points), subspace dimension $d/D$)?

<table>
<thead>
<tr>
<th>Function</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>[mu, Ud, Y] = r pca_r l s(X, d)</code></td>
<td>Finds the parameters of the PCA model $\mu$ and $U_d$ and the low-dimensional representation using re-weighted least squares with weights $w(e) = \frac{e^2}{\sigma^2 + e^2}$.</td>
</tr>
<tr>
<td><code>[L, E] = r pca(X, tau, 'method')</code></td>
<td>Solves the optimization problem $\min_{L,E} |L|<em>* + \lambda|E|</em>\ell$ subject to $X = L + E$ where $\ell = \ell_1$ or $\ell = \ell_{2,1}$ using the ADMM algorithm.</td>
</tr>
</tbody>
</table>

**Submission instructions.** Please follow the same instructions as in HW1.