Contents

(This doc has notes of topics from non-convex optimization which I might want to revisit. Also, has my attempts of the exercises given in the book "Non-Convex Optimization for Machine Learning". The points highlighted in Red are the ones I need to expand on).

1 General Points

- The solutions to the convex relaxation, e.g. LASSO relaxation for sparse linear regression, can be poor solutions to the original problem in general.
- Problem structures that allow non-convex approaches to avoid NP-Hardness results, are very similar to those that allow their convex relaxation counterparts to avoid distortions and a large relaxation gap.

1.1 Examples

1.1.1 Low Rank Matrix Completion

$$\begin{aligned} &\underset{X \in \mathbb{R}^{m \times n}}{\arg\min} \sum_{(i,j) \in [m] \times [n]} (X_{ij} - A_{ij})^2 \\ &\text{s.t. } rank(X) \leq r \end{aligned}$$

Convex objective but non-convex rank constraint.

Alternate Formulation Assuming A has at most r rank is equivalent to assuming A can be written as $A = UV^\top$ where $U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r}$. As $(UV^\top)_{i,j} = U_i^\top V_j$ where U_i is the i^{th} row of U and V_j is the j^{th} row of V, the formulation becomes

Formulation

$$\underset{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r}}{\arg\min} \sum_{(i,j) \in [m] \times [n]} \left(U_i^\top V_j - A_{ij}\right)^2$$

TODO: The objective is non-convex in (U, V)

2 Some Basic Results

3 My Attempts to Exercises

3.1 (Q) 2.1

Show strong smoothness does not imply convexity by constructing a non-convex function $f : \mathbb{R}^p \to \mathbb{R}$ that is 1-SS.

My Attempt $f(\mathbf{x}) = \frac{-1}{2}\|\mathbf{x}\|_2^2$.

$$f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle = \frac{-\|\mathbf{y} - \mathbf{x}\|_2^2}{2} \leq \frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2} \tag{1}$$

3.2 (Q) 2.3

Show that for any point $\mathbf{z} \notin \mathcal{B}_2(r)$, the projection onto the ball is $\frac{r}{\|\mathbf{z}\|_2}\mathbf{z}$

My Attempt **By Contradiction:** Let $\mathbf{z}' \neq \frac{r}{\|\mathbf{z}\|_2}\mathbf{z}$ be the projection.
$\frac{r}{\|\mathbf{z}\|_2}\mathbf{z} \in \mathcal{B}_2(r)$. From the definition of projection,

$$\|\mathbf{z}' - \mathbf{z}\| \leq \|\frac{r}{\|\mathbf{z}\|_2}\mathbf{z} - \mathbf{z}\| \implies \|\mathbf{z}' - \mathbf{z}\| \leq \|r - \|\mathbf{z}\|_2\|$$

From Triangle Inequality, $\|\mathbf{z}' - \mathbf{z}\| \geq \|r - \|\mathbf{z}\|\|$ as $\|\mathbf{z}'\| = r$. From the above two, $\|\mathbf{z}' - \mathbf{z}\| = \|r - \|\mathbf{z}\|\|$. But, by Triangle Inequality, $\|a - b\| = \|\|a\| - \|b\|\|$ iff $a = b$. This implies that $\mathbf{z}' = \mathbf{z}$ which can not happen as $\mathbf{z} \notin \mathcal{B}_2(r)$ and $\mathbf{z}' \in \mathcal{B}_2(r)$.

3.3 (Q) 2.5

Show that if $f : \mathbb{R}^p \to \mathbb{R}$ is a strongly convex function that is differentiable, then there is a unique $\mathbf{x}^* = \arg\min_{x \in \mathbb{R}^p} f(\mathbf{x})$

My Attempt **By Contradiction:** Let $\mathbf{x}^*, \mathbf{y}^*$ be the solutions of $\arg\min_{x \in \mathbb{R}^p} f(\mathbf{x})$ such that $\mathbf{x}^* \neq \mathbf{y}^*$. Then $\nabla f(\mathbf{y}^*) = \nabla f(\mathbf{x}^*) = 0$. From strong-convexity of f, we have that

$$f(\mathbf{y}^*) - f(\mathbf{x}^*) - \langle \nabla f(\mathbf{x}^*), \mathbf{y}^* - \mathbf{x}^* \rangle \geq \frac{\alpha}{2}\|\mathbf{x}^* - \mathbf{y}^*\|_2^2$$
$$f(\mathbf{x}^*) - f(\mathbf{y}^*) - \langle \nabla f(\mathbf{y}^*), \mathbf{x}^* - \mathbf{y}^* \rangle \geq \frac{\alpha}{2}\|\mathbf{y}^* - \mathbf{x}^*\|_2^2$$

Adding these, we get $0 \geq \|\mathbf{x}^* - \mathbf{y}^*\|_2^2 \implies \mathbf{x}^* = \mathbf{y}^*$.

3.4 (Q) 2.6

Show that $\mathcal{B}_0(s) \subset \mathbb{R}^p$ is non-convex for any $s < p$. What happens when $s = p$?

$$\mathcal{B}_0(s) = \{\|\mathbf{x}\|_0 \leq s\}$$

Let $\mathbf{x}, \mathbf{y} \in \mathcal{B}_0(s) \implies \|\mathbf{x}\|_0 \leq s, \|\mathbf{y}\|_0 \leq s$. As $s < p$, $\|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\|_0 \leq 2s - p$. Hence, $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$ for $\lambda \in (0, 1)$ need not belong to $\mathcal{B}_0(s)$ for any $s < p$. For $s = p$, $\mathcal{B}_0(s)$ will be convex.