Projection onto convex set

Yinan Huang

Nov 14, 2020

Contents

1 Introduction

In the study of best approximation problem and SVD, we find the concept of (orthogonal) projection onto subspaces plays an important role in understanding them. In general, we can generalize projection to convex set, which is defined in a natural way from projection onto subspace. It is extremely useful in convex optimization problem, for example, the best approximation problem. In that case, it turns out the solution set is convex and we can obtain a solution by project any vector onto the solution set.

2 Convex set and convex functional

We are going to introduce the concepts of convex sets and convex funtionals, which actually defines the convex optimization problem: minimize a convex functional on a closed convex set. It can be proved that an convex optimization problem always has solution. And it turns out the projection onto convex set is defined as a convex optimization problem, which we will see in the next section.

2.1 Convex set

Definition 2.1. *Let V be a vector space. A subset $A \subseteq V$ is said to be **convex**, if for any $\boldsymbol{x}, \boldsymbol{y} \in A$, for any $\lambda \in (0,1)$, we have*

$$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in A. \tag{1}$$

Basically, convexity means if we draw a line between two point in the set, every point on the line should also be in the set.

Theorem 2.2. *Intersection of any convex sets is also convex.*

Proof 2.3. *Let $A = \cap_i A_i$ and A_i are all convex. Let $\boldsymbol{x}, \boldsymbol{y} \in A$, then $\boldsymbol{x}, \boldsymbol{y} \in A_i$ for all i. Notice that for any $\lambda \in (0,1)$, $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in A_i$ because of convexity of A_i. So $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in A$ by definition. Thus A is convex.*

It is obvious that subspace is convex. Just as superplane is a common subspace in all subspaces, affine superplane is also important in convex sets.

Definition 2.4. *In vector space V, **hyperplane** A is a subset defined by $A = \{\boldsymbol{v} \in V | \langle \boldsymbol{v} | \boldsymbol{u_0} \rangle = 0\}$. $\boldsymbol{u_0}$ is called the normal vector, and all vectors in the hyperplane is orthogonal to $\boldsymbol{u_0}$. **Affine hyperplane** C is a subset defined by $C = \{\boldsymbol{v} \in V | \langle \boldsymbol{v} | \boldsymbol{u_0} \rangle = c\}$.*

Theorem 2.5. *Hyperplane and affine hyperplane are convex.*

Proof 2.6. *Hyperplane is convex since it is a subspace. For any $\boldsymbol{x}, \boldsymbol{y}$ in affine hyperplane $C = \{\boldsymbol{v} \in V | \langle \boldsymbol{v} | \boldsymbol{u_0} \rangle = c\}$, we have $\langle \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} | \boldsymbol{u_0} \rangle = \lambda c + (1-\lambda)c = c$, so C is convex.*

Affine hyperplane be seen as a hyperplane after displacement. Let $\boldsymbol{v_0}$ be a vector such that $\langle \boldsymbol{v_0} | \boldsymbol{u_0} \rangle = c$, then affine hyper plane is just $C = A + \boldsymbol{v_0}$ (this is where the word "affine" comes). In general, an affine set $A + \boldsymbol{v_0}$ of a convex set A is also convex, and thus was called affine convex set.

2.2 Convex functional

Definition 2.7. *A functional $f : V \to \mathbb{R}$ is called* **convex** *on convex set C, if for any $\boldsymbol{x}, \boldsymbol{y} \in C$ and any $\lambda \in (0,1)$, we have*

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}). \tag{2}$$

If equality holds only for $\boldsymbol{x} = \boldsymbol{y}$, then it is called strictly convex.

Note that to have a convex functional, we first require a convex set. Also note that a linear functional is convex, since equality holds always by the definition of linearity.

Theorem 2.8. *For a convex functional f on convex set C, its local minimum are all equal (if exists) and thus is the global minimum; If f is stricly convex and global minimum exists, then there is a unique $\boldsymbol{x}^* \in C$ that achives the global minimum.*

Proof 2.9. *Let $f(\boldsymbol{x_1})$ be a local minimum, and $f(\boldsymbol{x_2}) < f(\boldsymbol{x_1})$. By the definition of local minimum, there exists a $\delta > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x_1})$ for $\boldsymbol{x} \in B(\boldsymbol{x_1}, \delta)$. Consider $\lambda < \frac{\delta}{\|\boldsymbol{x_1} - \boldsymbol{x_2}\|}$ and $\boldsymbol{x} = (1-\lambda)\boldsymbol{x_1} + \lambda \boldsymbol{x_2}$, then $\boldsymbol{x}$ is just in the ball $B(\boldsymbol{x}, \delta)$ since*

$$\|\boldsymbol{x} - \boldsymbol{x_1}\| = \|\lambda \boldsymbol{x_2} - \lambda \boldsymbol{x_1}\| = \lambda(\|\boldsymbol{x_1} - \boldsymbol{x_2}\|) < \delta.$$

Then by definition of local minimum we should have $f(\boldsymbol{x}) \geq f(\boldsymbol{x_1})$. But the convexity says

$$f(\boldsymbol{x}) \leq \lambda f(\boldsymbol{x_2}) + (1-\lambda) f(\boldsymbol{x_1}) < f(\boldsymbol{x_1}). \tag{3}$$

So it is a contradiction. So we can only have $f(\boldsymbol{x_2}) \geq f(\boldsymbol{x_1})$ in order to maintain consistence. Thus local minimum $f(\boldsymbol{x_1})$ (if exists) are global minimum indeed. If f is stricly convex, then to avoid a contradiction, we must require $f(\boldsymbol{x_2}) > f(\boldsymbol{x_1})$, so there is at most one local minimum (and global minimum as well).

The theorem above essentialy shows the sigificant properties of a convex functional, that is, one find the global minimum as he find any local minimum. But we have not yet discussed about the existence. If convex set is **compact**, then the condition is strong enough to ensure the existence.

Theorem 2.10. *A convex functional $f : V \to \mathbb{R}$ on* **compact** *convex set C implies the existence of global minimum (though too strong).*

Proof 2.11. *Recall that compactness of a subset ensures the achivement of global minimum for any function.*

3 Projection onto convex set

3.1 Projection onto general convex set

Recall that (orthogonal) projection of vector $\boldsymbol{v}$ onto a subspace U implies that

$$\|P_U(\boldsymbol{v}) - \boldsymbol{v}\| = \min_{\boldsymbol{u} \in U} \|\boldsymbol{u} - \boldsymbol{v}\|. \tag{4}$$

We can generalize this to projection onto convex set.

Definition 3.1. *Let C be a convex set.* ***Projection*** *$P_C : V \to V$ of vector $\boldsymbol{v}$* ***onto*** *C is defined by:*

$$P_C(\boldsymbol{v}) \equiv \arg\min_{\boldsymbol{u} \in C} \|\boldsymbol{u} - \boldsymbol{v}\|. \tag{5}$$

Theorem 3.2. *Let V be Hilbert space and C be a **closed** convex set, then projection of $\boldsymbol{v} \in V$ onto C **exists** and is **unique**.*

Proof 3.3. *Let first prove existence. Let $d = \inf_{\boldsymbol{u} \in C} \|\boldsymbol{u} - \boldsymbol{v}\|$ be the infimum. There must exist a squence $\{\boldsymbol{u_i}\} \subset C$ that converges to d:*

$$\left\| \lim_{n\to\infty} \boldsymbol{u_n} - \boldsymbol{v} \right\| = \lim_{n\to\infty} \|\boldsymbol{u_n} - \boldsymbol{v}\| = d. \tag{6}$$

Notice that C is complete since V is complete. So we are going to show $\boldsymbol{u}_n$ is a cauchy sequence in order to show $\boldsymbol{u}_\infty \in C$. Consider

$$\begin{aligned} \|\boldsymbol{u}_n - \boldsymbol{u_m}\|^2 &= \|\boldsymbol{u_n} - \boldsymbol{v} - (\boldsymbol{u_m} - \boldsymbol{v})\|^2 = 2\|\boldsymbol{u_n} - \boldsymbol{v}\|^2 + 2\|\boldsymbol{u_m} - \boldsymbol{v}\|^2 - \|\boldsymbol{u_n} - \boldsymbol{v} + \boldsymbol{u_m} - \boldsymbol{v}\|^2 \\ &= 2\|\boldsymbol{u_n} - \boldsymbol{v}\|^2 + 2\|\boldsymbol{u_m} - \boldsymbol{v}\|^2 + 4\left\|\boldsymbol{v} - \frac{\boldsymbol{u_n} + \boldsymbol{u_m}}{2}\right\|^2 \\ &\leq 2\|\boldsymbol{u_n} - \boldsymbol{v}\|^2 + 2\|\boldsymbol{u_m} - \boldsymbol{v}\|^2 + 4d^2. \end{aligned} \tag{7}$$

In the last step, we use the fact that $(\boldsymbol{u_n} + \boldsymbol{u_m})/2 \in C$ and the convexity of C. Note that if $n, m \to \infty$, then the RHS is zero. So $\{\boldsymbol{u_n}\}$ is cauchy, and thus the limit exists.

Uniqueness can be seen from the fact that this is a convex optimization problem. Since $\|\cdot\|$ is strictly convex on C, so its global minimum is unique once exists.

Recall that in orthogonal projection of $\boldsymbol{v}$ onto subspaces U, we have $\langle \boldsymbol{v} - \boldsymbol{u}^* | \boldsymbol{u} \rangle = 0$ for any $\boldsymbol{u} \in U$ with $P_U(\boldsymbol{v}) = \boldsymbol{u}^*$. This projection theorem is generalized to projection onto convex sets.

Theorem 3.4. *Let C be a convex set and $P_C(\boldsymbol{v}) = \boldsymbol{u}^*$ is a projection onto C. Then for each $\boldsymbol{u} \in C$, we have*

$$\operatorname{Re} \langle \boldsymbol{v} - \boldsymbol{u}^* | \boldsymbol{u} - \boldsymbol{u}^* \rangle \leq 0. \tag{8}$$

Proof 3.5. *A rigorous proof can be found in ant text book, and we will give a picture here. If you project any $\boldsymbol{v}$ onto C and get $P_C(\boldsymbol{v}) = \boldsymbol{u}^*$, then for any other $\boldsymbol{u} \in C$, it is easy to see that the angle between $\boldsymbol{v} - \boldsymbol{u}^*$ and $\boldsymbol{u} - \boldsymbol{u}^*$ is greater than 90 degree (which is exactly 90 degree for subspaces projection), so the inner product must be less than 0.*

Theorem 3.6 (Projection onto affine convex set)**.** *Let C be a convex set and $A = C + \boldsymbol{u_0}$ be an affine convex set. Then projection onto A is given by*

$$P_A(\boldsymbol{v}) = \boldsymbol{u_0} + P_C(\boldsymbol{v} - \boldsymbol{u_0}). \tag{9}$$

Proof 3.7. *Let $\boldsymbol{v} \in V$ and by definition we have*

$$\|\boldsymbol{v} - P_A(\boldsymbol{v})\| \leq \|\boldsymbol{v} - \boldsymbol{u}\|$$

for all $\boldsymbol{u} \in A$. Notice that since $A = C + \boldsymbol{u_0}$, we can recast it as

$$\|\boldsymbol{v} - P_A(\boldsymbol{v})\| \leq \|\boldsymbol{v} - \boldsymbol{u}\| = \|\boldsymbol{v} - (\boldsymbol{u} + \boldsymbol{u_0})\| = \|\boldsymbol{v} - \boldsymbol{u_0} - \boldsymbol{u}\|$$

for all $\boldsymbol{u} \in C$. So if we see $\boldsymbol{v} - \boldsymbol{u_0}$ as a random vector, then the LHS is exactly the definition of $\|\boldsymbol{v} - \boldsymbol{u_0} - P_C(\boldsymbol{v} - \boldsymbol{u_0})\|$, so

$$P_A(\boldsymbol{v}) = \boldsymbol{u_0} + P_C(\boldsymbol{v} - \boldsymbol{u_0}).$$

3.2 Projection onto superplane

Superplane plays an important role in convex sets, so we would like to study the projection onto hyperplanes. Notice that a hyperplane $C = \{\boldsymbol{v} \in V | \langle \boldsymbol{v} | \boldsymbol{u_0} \rangle = 0\}$ is a orthogonal complement of 1D space $\{c\boldsymbol{u_0} \in V | c \in F\}$. So the projection onto C is simply the error vector of projection onto $\{c\boldsymbol{u_0} \in V | c \in F\}$:

$$P_A(\boldsymbol{v}) = \boldsymbol{v} - \frac{\langle \boldsymbol{v} | \boldsymbol{u_0} \rangle \, \boldsymbol{u_0}}{\|\boldsymbol{u_0}\|^2}. \tag{10}$$

Applying the thorem of projection onto affine convex sets, we get the projection onto affine hyperplane $A + \boldsymbol{w} = \{\boldsymbol{v} \in V | \langle \boldsymbol{v} | \boldsymbol{u_0} \rangle = \langle \boldsymbol{w} | \boldsymbol{u_0} \rangle\}$ is

$$P_{A+\boldsymbol{w}}(\boldsymbol{v}) = \boldsymbol{v} - \frac{(\langle \boldsymbol{v} | \boldsymbol{u_0} \rangle - c) \, \boldsymbol{u_0}}{\|\boldsymbol{u_0}\|^2} \tag{11}$$

with $c = \langle \boldsymbol{w} | \boldsymbol{u_0} \rangle$.

3.3 Seperating hyperplane theorem

It turns out given a convex set C and any $\boldsymbol{v} \notin C$, we can always find a hyperplane $A = \{\boldsymbol{w} \in V | \langle \boldsymbol{w} | \boldsymbol{u_0} \rangle = c\}$ such that $\boldsymbol{v}$ is in one side of A and C is in another side.

Theorem 3.8 (Seperating hyperplane theorem)**.** *Let C be a convex set and $\boldsymbol{v} \in V$ be a vector that is not in C. There exists an affine hyperplane $A = \{\boldsymbol{w} \in V | \langle \boldsymbol{w} | \boldsymbol{u_0} \rangle = c\}$ such that*

$$\begin{aligned} &\langle \boldsymbol{v} | \boldsymbol{u_0} \rangle > c \\ &\langle \boldsymbol{w} | \boldsymbol{u_0} \rangle \leq c \quad \textit{for all } \boldsymbol{w} \in A. \end{aligned} \tag{12}$$

Proof 3.9. *Intuitively, the hyperplane is simply the tangent plane that includes $P_C(\boldsymbol{v})$ with normal vector $\boldsymbol{v} - P_C(\boldsymbol{v})$. Concretely, let us define the affine hyperplane by*

$$A \equiv \{\boldsymbol{w} \in V | \langle \boldsymbol{w} | \boldsymbol{v} - \boldsymbol{v}^* \rangle = \langle \boldsymbol{v}^* | \boldsymbol{v} - \boldsymbol{v}^* \rangle\} \tag{13}$$

with $\boldsymbol{v}^* \equiv P_C(\boldsymbol{v})$. Let us consider

$$\langle \boldsymbol{v} | \boldsymbol{v} - \boldsymbol{v}^* \rangle - \langle \boldsymbol{v}^* | \boldsymbol{v} - \boldsymbol{v}^* \rangle = \|\boldsymbol{v} - \boldsymbol{v}^*\|^2 > 0, \tag{14}$$

so $\boldsymbol{v}$ lies in the positive direction of the hyperplane (in the direction of the normal vector). Then consider for $\boldsymbol{u} \in C$,

$$\langle \boldsymbol{u} | \boldsymbol{v} - \boldsymbol{v}^* \rangle - \langle \boldsymbol{v}^* | \boldsymbol{v} - \boldsymbol{v}^* \rangle \leq 0 \tag{15}$$

due to Theorem 3.4. So the whole convex set is in the another direction of the hyperplane. In this sense, Theorem 3.4 is saying a simple fact that the hyperplane induced by $P_C(\boldsymbol{v}) = \boldsymbol{v}^*$ seperates the convex set in one side, which is obvious visually.

4 Alternating projection

Alternating projection is a technics to approximate projection onto a complex set by projecting onto simple subsets alternatively.

Theorem 4.1. *Let $A = \cap_i^N A_i$ with A_i being closed sets. Then consider the alternating projection:*

$$\boldsymbol{v_{n+1}} = P_{A_{(n \bmod N)}}(\boldsymbol{v_n}) \tag{16}$$

with random initial $\boldsymbol{v_0}$. Then if $\boldsymbol{v_\infty}$ exists, it will converge to some point in A.

Alternating projection is useful when the subsets A_i are simple and while A is complicated. For example, consider linear equation

$$A\boldsymbol{x} = \boldsymbol{b}. \tag{17}$$

Our solution set A is

$$A = \{\boldsymbol{x} \in V | \, \langle \boldsymbol{x} | \boldsymbol{a_i} \rangle = b_i, \text{for all } i\} = \cap_i \{\boldsymbol{x} \in V | \, \langle \boldsymbol{x} | \boldsymbol{a_i} \rangle = b_i\} \equiv \cap_i A_i, \tag{18}$$

where $\boldsymbol{a_i}$ is the i^{th} row vector of matrix A. Notice that A_i, one of the equations, is simple and represents an affine hyperplane. So we use alternating projection to projection any $\boldsymbol{v_0}$ onto A_i, which is simple, and iteratively we can find one solution we want:

$$\boldsymbol{v}_{n+1} = P_{A_{(n \bmod N)}}(\boldsymbol{v_n}) = \boldsymbol{v_n} - \frac{(\langle \boldsymbol{v_n} | \boldsymbol{a_i} \rangle - b_i)\, \boldsymbol{a_i}}{\|\boldsymbol{a_i}\|^2} \tag{19}$$

Note that if we begin with $\boldsymbol{v_0} = 0$, then hopefully we can obtain a minimum norm solution. For general convex set, it is not surprising that alternating projection is also useful, since we can see a convex set as infinite intersection of hyperplanes, and projection onto them are fairly simple.

5 Conclusion

We introduced the concepts of convex set and convex functional, which are the fundamental ideas in convex optimization. Then we discuss about projection onto convex set, and it is a common topic in convex optimization: how to minimize the norm. Finally we talk about alternating projection, which is a method to projection onto complicated convex set by calculating projection onto simple convex subset (usually affine hyperplanes).