How can we project to an arbitrary norm ball?

1 Major Lessons

I think the general way of doing this is to go for proximal operators. Proximal operator of a closed proper convex function $f(\cdot)$ is defined as:

$$\mathbf{prox}_f(\mathbf{v}) = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{v} - \mathbf{x}\|_2^2 + f(\mathbf{x}) \tag{1}$$

Note that the objective of the above optimization is strongly convex, so there is a unique minimizer for it.

1.1 Generalizing Projection

Proximal operator is a generalization of projection in the following sense. Remember the indicator of a set $\mathcal{S}$, i.e.,

$$\mathbb{1}_{\mathcal{S}}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in \mathcal{S} \\ \infty & \mathbf{x} \notin \mathcal{S} \end{cases} \tag{2}$$

Then the proximal operator reduces to Euclidian projection:

$$\mathbf{prox}_f(\mathbf{v}) = \underset{\mathbf{x}\in\mathcal{S}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{v} - \mathbf{x}\|_2^2 = \Pi_{\mathcal{S}}(\mathbf{v}) \tag{3}$$

1.2 Generalizing Orthogonal Decomposition

Let $\mathcal{S}$ and $\mathcal{S}^\perp$ be two orthogonal subspaces. Then for any vector $\mathbf{v}$ we have:

$$\mathbf{v} = \Pi_{\mathcal{S}}(\mathbf{v}) + \Pi_{\mathcal{S}^\perp}(\mathbf{v}) \tag{4}$$

We have the same decomposition for closed proper convex f and its convex conjugate $f^*(\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathbf{x}^T\mathbf{y} - f(\mathbf{x})$:

$$\mathbf{v} = \mathbf{prox}_f(\mathbf{v}) + \mathbf{prox}_{f^*}(\mathbf{v}) \tag{5}$$

The above result is known as **Moreau Decomposition**.

Proof: For the $\mathbf{u} = \mathbf{prox}_f(\mathbf{v})$, from the optimality condition we should have:

$$\mathbf{v} - \mathbf{u} \in \partial f(\mathbf{u}) \iff \mathbf{u} \in \partial f^*(\mathbf{v} - \mathbf{u}) \iff (\mathbf{v} - \mathbf{u}) = \mathbf{prox}_{f^*}(\mathbf{u}) \tag{6}$$

Above, we used a useful property which relates subgradient of f and f^*:

$$\mathbf{u} \in \partial f(\mathbf{v}) \iff \mathbf{v} \in \partial f^*(\mathbf{u}) \iff f^*(\mathbf{u}) + f(\mathbf{v}) = \mathbf{u}^T\mathbf{v} \tag{7}$$

1.3 Moreau Decomposotion for a Norm

Here we first need to following for an arbitrary norm $f(\mathbf{x}) = \|\mathbf{x}\|$:

$$f^*(\mathbf{y}) = \mathbb{1}_{\|\cdot\|_* \leq 1}(\mathbf{y}) \tag{8}$$

where $\|\cdot\|_*$ is the daul norm. In other words, the convex conjugate of a norm is the indicator function of its dual norm.

From the above equality and the Moreau decomposition we get:

$$\begin{aligned} \mathbf{v} &= \mathbf{prox}_{\|\cdot\|} + \mathbf{prox}_{\mathbb{1}_{\|\cdot\|_* \leq 1}(\cdot)} \\ &= \mathbf{prox}_{\|\cdot\|} + \Pi_{\mathbb{1}_{\|\cdot\|_* \leq 1}}(\mathbf{v}) \end{aligned}$$

Above equality gives us a way to switch between proximal operator **prox** and projection Π.