Recitation 21

Thursday November 21, 2024

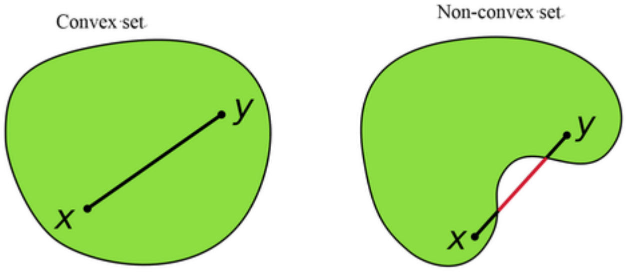
1 Recap

1.1 Convex Set

A set C is *convex* if for all $x_1, x_2 \in C$ and all $t \in [0, 1]$,

$$tx_1 + (1-t)x_2 \in C.$$

In other words, C contains the line segment between any two points in C.



The projection of a point y onto a set C is given by the closest point in the set to y

$$\mathrm{Proj}_C(y) = \arg\min_{z \in C} \|z - y\|.$$

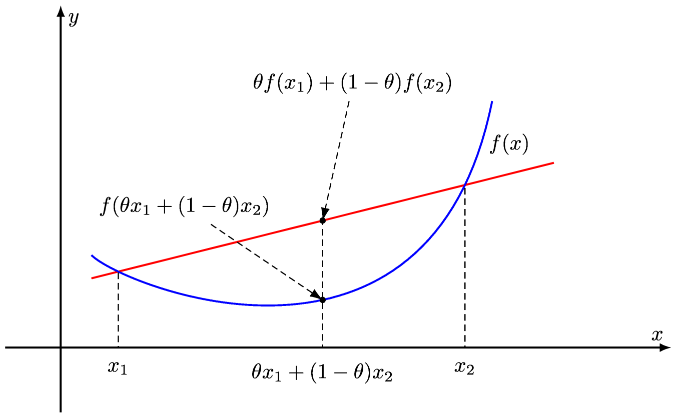
For example, if C is a box given by $[0, 1]^2$ and $y = (1.5, 0.5)$, then $\mathrm{Proj}_C(y) = (1, 0.5)$.

1.2 Convex Function

Let $X \subseteq \mathbb{R}^n$ be a convex set. A function $f : X \to \mathbb{R}$ is *convex* if for any pair of points $x_1, x_2 \in X$ and any $t \in [0, 1]$, we have

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2),$$

i.e., the graph of $f(x)$ lies below the line segment connecting $f(x_1)$ and $f(x_2)$.



1.3 Taylor Expansion for Differentiable Functions

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the **first-order Taylor expansion** around a point x_k is:

$$f(x) \approx f(x_k) + \nabla f(x_k)^T (x - x_k).$$

This linear approximation estimates the function value near x_k. If f is twice differentiable, the **second-order Taylor expansion** includes the Hessian matrix $\nabla^2 f(x_k)$:

$$f(x) \approx f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).$$

This quadratic approximation provides a more accurate estimate near x_k and is used in methods like Newton's method.

1.4 Projected Gradient Descent

We can extend the quadratic minimization framework to constrained optimization by projecting the gradient step back into a feasible set C. When minimizing $g(x)$ subject to $x \in C$, we have:

$$\min_{x \in C} \left[f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2\alpha} \|x - x_k\|^2 \right].$$

This is equivalent to:

1. Taking an unconstrained gradient step: $y = x_k - \alpha \nabla f(x_k)$.
2. Projecting y back onto C: $x_{k+1} = \mathrm{Proj}_C(y) = \arg\min_{z \in C} \|z - y\|$.

Thus, the update rule for projected gradient descent is:

$$x_{k+1} = \mathrm{Proj}_C \left(x_k - \alpha \nabla f(x_k) \right).$$

2 Exercises

1. **Projection onto Convex Sets** Given the following sets, find an expression for $\text{Proj}_C(y)$ for any point y in the same space:

   (a) **Box**: $C = [0, 1]^3$

   (b) **Linear Subspace**: $C = \{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 0\}$

   (c) **Affine Plane**: $C = \{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = -1\}$

   (d) **Half-Space**: $C = \{x \in \mathbb{R}^2 \mid x_1 + x_2 \leq 4\}$

   (e) **Ball**: $C = \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 2\}$

2. In this problem, we are going to express Gradient Descent as a quadratic minimization. We want to minimize a quadratic approximation (a first-order Taylor expansion plus a quadratic penalty term). Consider a differentiable function $f(x)$:

$$g(x) = \nabla f(x_k)^T (x - x_k) + \frac{1}{2\alpha}\|x - x_k\|^2.$$

   (a) Show that minimizing the function $g(x)$ with respect to x yields the gradient descent update $x_{k+1} = x_k - \alpha \nabla f(x_k)$.

   (b) Show that this minimization is equivalent to solving the unconstrained quadratic minimization problem:
$$\min_x \left[\frac{1}{2}x^T A x + b^T x\right],$$
   where $A = \frac{1}{\alpha}I$ and $b = -\frac{1}{\alpha}x_k - \nabla f(x_k)$.

   (c) Verify that the solution to the quadratic minimization problem in part (b) is the same as the gradient descent update.

3. **Projected Gradient Descent.** For this problem, we are interested in minimizing the quadratic approximation as similar to problem (2):

$$g(x) = \nabla f(x_k)^T (x - x_k) + \frac{1}{2\alpha}\|x - x_k\|^2.$$

   However, we are looking for solutions within the set C:

$$\min_{x \in C} \left[\nabla f(x_k)^T (x - x_k) + \frac{1}{2\alpha}\|x - x_k\|^2\right].$$

   (a) Show that the solution x_{k+1} to this problem is given by:
$$x_{k+1} = \text{Proj}_C\left(x_k - \alpha \nabla f(x_k)\right).$$
   Hint: Use the definition of projection onto a set C.

   (b) Explain why this procedure can be interpreted as taking a gradient step followed by a projection onto C.

3 Solutions

1. **Projection onto Different Convex Sets**

   (a) **Box Constraint** $C = [0,1]^3$:
   For any $y = (y_1, y_2, y_3) \in \mathbb{R}^3$, the projection onto C is given by:

   $$x_i = \min\left(\max(y_i, 0), 1\right), \quad \forall i = 1, 2, 3.$$

   That is, each component is clipped to the interval $[0,1]$.

   (b) **Plane Constraint** $C = \{x \in \mathbb{R}^3 \mid a^T x = b\}$, where $a = (1,1,1)$ and $b = -1$:
   The projection onto the plane is:

   $$x = y - \frac{a^T y - b}{\|a\|^2} a$$

   (c) **Half-Space** $C = \{x \in \mathbb{R}^2 \mid a^T x \leq b\}$, where $a = (1,1)$ and $b = 4$:
   The projection is:

   $$x = \begin{cases} y, & \text{if } a^T y \leq b, \\ y - \frac{a^T y - b}{\|a\|^2} a, & \text{if } a^T y > b. \end{cases}$$

   That is, if y is already in the half-space, it remains unchanged; otherwise, it is projected onto the boundary.

   (d) **Ball** $C = \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq r\}$, with radius $r = 2$:
   The projection is:

   $$x = \begin{cases} y, & \text{if } \|y\|_2 \leq r, \\ r\dfrac{y}{\|y\|_2}, & \text{if } \|y\|_2 > r. \end{cases}$$

   That is, y is scaled back to lie on the boundary of the ball if it is outside.

2. **Gradient Descent as Quadratic Minimization**

   (a) To minimize $g(x)$, take the gradient with respect to x and set it to zero:

   $$\nabla g(x) = \nabla f(x_k) + \frac{1}{\alpha}(x - x_k) = 0.$$

   Solving for x:

   $$x = x_k - \alpha \nabla f(x_k).$$

   This is the gradient descent update rule.

   (b) The function $g(x)$ can be rewritten as:

   $$\begin{aligned} g(x) &= \frac{1}{2\alpha}\|x - x_k\|^2 + \nabla f(x_k)^T (x - x_k) \\ &= \frac{1}{2\alpha}(x - x_k)^T (x - x_k) + \nabla f(x_k)^T (x - x_k). \end{aligned}$$

   Expanding and rearranging terms:

   $$g(x) = \frac{1}{2\alpha} x^T x - \left(\frac{1}{\alpha} x_k + \nabla f(x_k)\right)^T x + \text{constant}.$$

Therefore, minimizing $g(x)$ is equivalent to solving:

$$\min_x \left[\frac{1}{2}x^T Ax + b^T x\right],$$

where $A = \frac{1}{\alpha}I$ and $b = -\left(\frac{1}{\alpha}x_k + \nabla f(x_k)\right)$.

(c) The solution to the quadratic minimization problem is:

$$x = -A^{-1}b = -\left(\frac{1}{\alpha}I\right)^{-1}\left(-\left(\frac{1}{\alpha}x_k + \nabla f(x_k)\right)\right) = x_k - \alpha\nabla f(x_k).$$

This matches the gradient descent update.

3. **Projected Gradient Descent**

(a) The constrained minimization problem can be rewritten as:

$$x_{k+1} = \arg\min_{x\in C}\left[\nabla f(x_k)^T(x - x_k) + \frac{1}{2\alpha}\|x - x_k\|^2\right].$$

Completing the square and ignoring constants (since they do not affect the minimization), we have:

$$\min_{x\in C}\left[\frac{1}{2\alpha}\|x - x_k + \alpha\nabla f(x_k)\|^2\right].$$

Therefore:

$$x_{k+1} = \arg\min_{x\in C}\|x - (x_k - \alpha\nabla f(x_k))\|^2 = \mathrm{Proj}_C\left(x_k - \alpha\nabla f(x_k)\right).$$

(b) This procedure first takes an unconstrained gradient step $y = x_k - \alpha\nabla f(x_k)$ as if there were no constraints. Then, since y might not be in C, we project y back onto C to obtain x_{k+1}. This ensures that the iterates remain feasible with respect to the constraints defined by C.