Projected Gradient Descent

Axel Böhm

October 16, 2021

Constrained Optimization

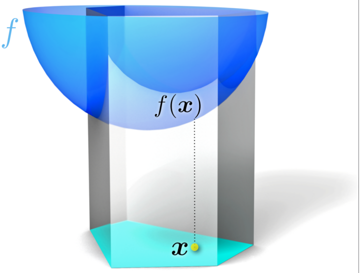
Constrained optimization problem

$$\text{minimize } f(x)$$

$$\text{subject to } x \in C$$

How to solve them

- Project onto C
- transform to *unconstrained problem*



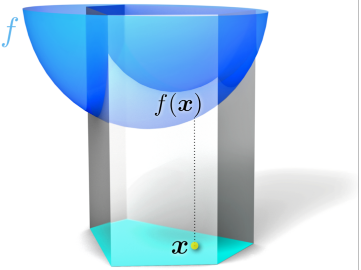
Constrained Optimization

Constrained optimization problem

$$\text{minimize } f(x)$$
$$\text{subject to } x \in C$$

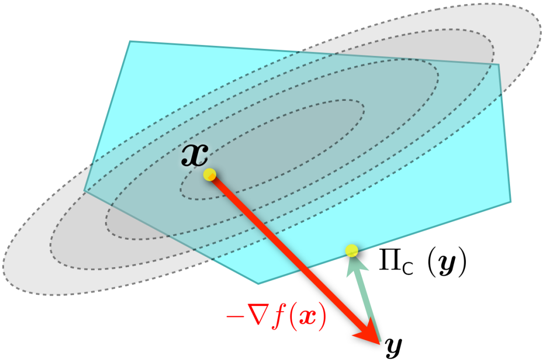
We will focus on:

◇ **Projected Gradient Descent**



Projected Gradient Descent

Idea: After every step project back onto the set:
$\Pi_C(x) := \arg\min_{y \in C} \|y - x\|$.



Projected subgradient method

$$\text{(constrained setting)} \quad \min_{x \in C} f(x)$$

Algorithm Projected subgradient method

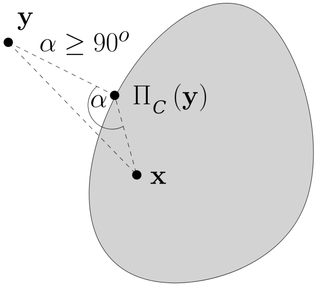
1: **for** $k = 0, 1, \ldots$ **do**
2: Pick $g_k \in \partial f(x_k)$
3: $y_{k+1} = x_k - \alpha g_k$
4: $x_{k+1} = \Pi_C(y_{k+1})$

Properties of the Projection

Fact

Let $C \subseteq \mathbb{R}^d$ be closed and convex, $x \in C$ and $y \in \mathbb{R}^d$. Then

- $\langle x - \Pi_C(y), y - \Pi_C(y) \rangle \leq 0$
- $\|x - \Pi_C(y)\|^2 + \|y - \Pi_c(y)\|^2 \leq \|y - x\|^2$



Properties of the Projection

Fact

Let $C \subseteq \mathbb{R}^d$ be closed and convex, $x \in C$ and $y \in \mathbb{R}^d$.Then

- $\langle x - \Pi_C(y), y - \Pi_C(y) \rangle \leq 0$
- $\|x - \Pi_C(y)\|^2 + \|y - \Pi_c(y)\|^2 \leq \|y - x\|^2$

Proof.

Since $\Pi_C(x)$ is the minimizer of a differentiable convex function $d_x(y) = \frac{1}{2}\|y - x\|^2$ over C, by the **first-order optimality condition**

$$
\begin{aligned}
0 &\leq \langle \nabla d_x(\Pi_C(x)), y - \Pi_C(x) \rangle \\
&= \langle \Pi_C(x) - x, y - \Pi_C(x) \rangle
\end{aligned}
$$

□

Results for projected GD

For **closed**, **convex** set $C \subset \mathbb{R}^d$ same number of gradient steps.

- Lipschitz convex function over C: $\mathcal{O}(\epsilon^{-2})$ steps
- Smooth convex function over C: $\mathcal{O}(\epsilon^{-1})$ steps
- Smooth and strongly convex over C: $\mathcal{O}(\log(\epsilon^{-1}))$

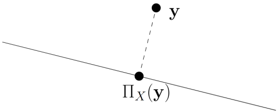
But:

- Each step requires a projection onto C
- May or may not be easy to compute

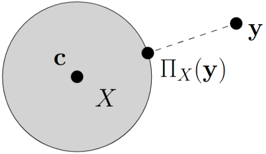
The projection step: $\Pi_C(x) := \arg\min_{y \in C} \|y - x\|$

Computing $\Pi_C(x)$ is an optimization problem itself.
Efficient in relevant cases:

- Box constraints: $C = [a_1, b_1] \times \cdots \times [a_d, b_d]$
- Affine subspace (requires solution of system of linear equations)



- Projection onto ball with center c



Convergence analysis

Proof.

We can deduce the exact same inequality as before

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|\Pi_C(x_k - \alpha g_k) - \Pi_C(x^*)\|^2 \\
&\leq \|x_k - \alpha g_k - x^*\|^2 \\
&= \|x_k - x^*\|^2 + 2\alpha\langle g_k, x^* - x_k\rangle + \alpha^2\|g_k\|^2 \\
&\leq \|x_k - x^*\|^2 + 2\alpha(f^* - f(x_k)) + \alpha^2\|g_k\|^2.
\end{aligned}$$

Continue the proof as in the unconstrained setting. □

Composite minimization problem

Consider objective function composed as

$$f(x) = g(x) + h(x)$$

where

- g is nice
- h is simple

typically we mean nice means smooth. Relevant if h is not differentiable. Most notably: Lasso

Idea

Classical gradient step for g:

$$x_{k+1} = \arg\min_x g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{1}{2\alpha}\|x - x_k\|^2$$

Now, for $f = g + h$ we keep this for g and add h unmodified:

$$\begin{aligned} x_{k+1} &= \arg\min_x g(x_k) + \langle \nabla g(x_k), x - x_k \rangle + \frac{1}{2\alpha}\|x - x_k\|^2 + h(x) \\ &= \arg\min_x \frac{1}{2\alpha}\|x - (x_k - \alpha \nabla g(x_k))\|^2 + h(x) \end{aligned}$$

The proximal gradient algorithm

An iteration is defined as

$$x_{k+1} = \text{prox}_{\alpha h}\left(x_k - \alpha \nabla g(x_k)\right)$$

where the proximal mapping for a function h and parameter α is defined as

$$\text{prox}_{\alpha h}(x) = \arg\min_{y \in \mathbb{R}^d} \left\{ h(y) + \frac{1}{2\alpha} \|y - x\|^2 \right\}.$$

A generalization of (projected) GD

- $h \equiv 0$ recovers gradient descent.
- $h = \chi_C$ recovers projected gradient descent We call χ_C the **indicator function** of C

$$\begin{aligned} \chi_C : \mathbb{R}^d &\to \mathbb{R} \cup +\infty \\ x &\mapsto \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Proximal mapping becomes

$$\text{prox}_{\alpha h}(x) = \arg\min_{y \in \mathbb{R}^d} \left\{ \chi_C(y) + \frac{1}{2\alpha}\|y - x\|^2 \right\} = \arg\min_{y \in C}\{\|y - x\|^2\}.$$

Convergence

Same complexity as GD or projected GD,
if we can compute the proximal mapping!