Convex Optimization Examples

Jaime Tenorio

April 2023

Convex Optimization

A *convex optimization problem* can be stated as

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \qquad i = 1, \ldots, m \\
& a_i^T x = b_i, \qquad i = 1, \ldots, p,
\end{aligned}
\tag{1}
$$

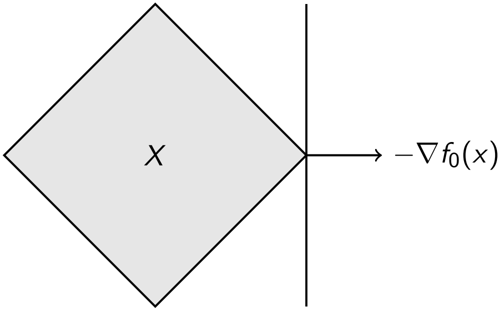
where the function $f_0(x)$ and the inequality constraints $f_i(x)$ are convex, and the equality constraint must be affine.
The feasible set is convex because it is the intersection of of the domain of the constraints, which are convex.

$$
D = \bigcap_{i}^{m} \mathbf{dom} f_i \tag{2}
$$

Suppose we have a convex optimization problem where the objective function f_0 is differentiable. Let X be the feasible set of the problem. Then $x \in X$ is optimal if and only if:

$$\nabla f_0(x)^T(y - x) \geq 0 \quad \forall y \in X \tag{3}$$

Geometrically, this means that the gradient of the objective function at x points in the direction of the feasible set, in other words, it is a supporting hyperplane to X.



Unconstrained Problems

Suppose we have no constraint inequalities f_i and the objective function f_0 is differentiable. Then $x \in X$ is optimal if and only if:

$$\nabla f_0(x) = 0 \tag{4}$$

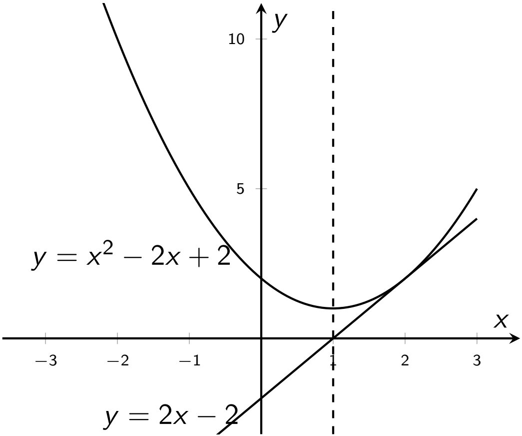
Take as an example the *unconstrained quadratic optimization*:

$$\text{minimize} \quad f_0(x) = \frac{1}{2} x^T P x + q^T x + r \tag{5}$$

where $P \in \mathbf{S}_+^n$. The optimality condition gives:

$$\nabla f_0(x) = Px + q = 0 \tag{6}$$

We can visualize this in two dimensions in the following figure:



Equality Constraints

If we have only equality constraints

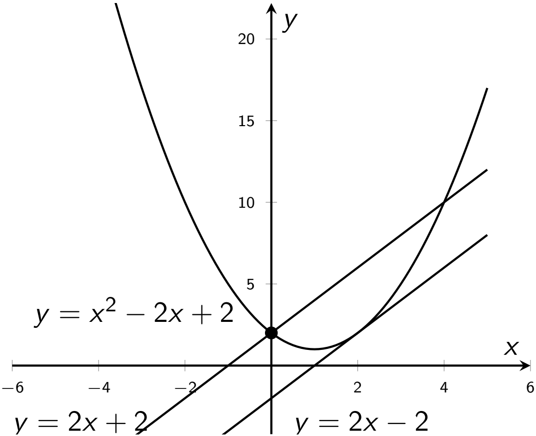
$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{7}$$

the feasible set is affine. The optimality condition for $x \in X$ is

$$\exists \nu \in \mathbb{R}^p \quad s.t. \quad \nabla f_0(x) + A^T \nu = 0 \quad \text{and} \quad Ax = b \tag{8}$$

which is the Lagrange multiplier optimality condition, also known as the Karush-Kuhn-Tucker (KKT) condition.

We can visualize this in two dimensions in the following figure:



Nonnegative Orthant

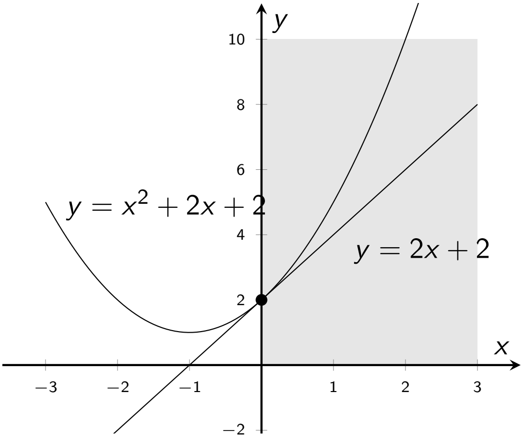
If we consider a problem like

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & x \succeq 0 \end{aligned} \tag{9}$$

where the only constraint is the nonnegativity of x. The feasible set is the nonnegative orthant. The optimality condition for $x \succeq 0, x \in X$ is

$$\begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases} \tag{10}$$

We can visualize this in two dimensions in the following figure:



Equivalent Problems

We can obtain **equivalent** convex problems using transformations that preserve convexity.

Original		Equivalent	
minimize subject to	$f_0(x)$ $f_i(x) \leq 0$ $Ax = b$	minimize (over z) subject to $Ax + b \iff x = Fz + x_0$	$f_0(Fz + x_0)$ $f_i(Fz + x_0)$
minimize subject to	$f_0(A_0x + b_0)$ $f_i(A_ix + b_i) \leq 0$	minimize (over x, y_i) subject to	$f_0(y_0)$ $f_i(y_i) \leq 0$ $y_i = A_ix + b_i$
minimize subject to	$f_0(x)$ $a_i^T x \leq b_i$	minimize (over x, s) subject to	$f_0(x)$ $a_i^T x + s_i = b_i$ $s_i \geq 0$
minimize subject to	$f_0(x)$ $f_i(x) \leq 0$ $a_i^T x = b_i$	minimize (over x, t) subject to	t $f_0(x) - t \leq 0$ $f_i(x) \leq 0$ $Ax = b$
minimize subject to	$f_0(x_1, x_2)$ $f_i(x_1) \leq 0$	minimize subject to $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$	$\tilde{f}_0(x_1)$ $f_i(x_1) \leq 0$

In the table we can see equivalent problems for the following transformations:

1. Eliminating equality constraints
2. Introducing equality constraints
3. Introducing slack variables for linear inequality constraints
4. Epigraph form of the standard convex problem
5. Minimizing over some variables

Neural Network Training

Training a neural network can be posed as a convex optimization problem.

$$\text{minimize} \quad L(y; \hat{y}) \tag{11}$$

where $L(y; \hat{y})$ is the loss function, y is the ground truth, and $\hat{y}$ is the prediction. The prediction is given by the neural network, which is a function of the weights W and the input x. If we consider a network of one layer, the prediction is given by

$$\hat{y} = \sigma(Wx) \tag{12}$$

where σ is the activation function. For this problem to be convex we need the loss function to be convex. We can consider the following loss functions:

- **Mean squared error (MSE)**: $L(y; \hat{y}) = \frac{1}{2}(y - \hat{y})^2$
- **Cross-entropy loss**: $L(y; \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$
- **Hinge loss**: $L(y; \hat{y}) = \max(0, 1 - y\hat{y})$

For example, YOLO is a convolutional neural network for object detection. It is trained using the following loss function

$$
\begin{aligned}
L = &= \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{i=0}^{B} \mathbf{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
&+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{i=0}^{B} \mathbf{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\
&+ \sum_{i=0}^{S^2} \sum_{i=0}^{B} \mathbf{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{i=0}^{B} \mathbf{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbf{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}
\tag{13}
$$

where $\hat{x}, \hat{y}, \hat{w}, \hat{h}, \hat{c}$ are the output from the convolutional neural network. This is a modified mean squared error loss function, which is convex.

Usually, the optimization of a neural network is not a convex problem.