21-259 Summer II 2017: Notes on optimization with multiple constraints

Let us review the methods of optimizing f we have already seen:

- Unconstrained optimization: without any constraints, the extrema of f will occur at critical points, so by setting $\nabla f = \vec{0}$, we will locate all extrema of f
- Optimization with equality constraint $g(\vec{x}) = 0$: under certain assumptions, the method of Lagrange multipliers says that all constrained extrema will occur where $\nabla f = \lambda \nabla g$ for some λ
- Optimization with inequality constraint $g(\vec{x}) \leq 0$: The extrema will either occur in the interior where $g(\vec{x}) < 0$ - in which case it will be a critical point of f - or on the boundary $g(\vec{x}) = 0$ - in which case it can be solved using Lagrangian multipliers.

We have also seen that in the case of optimizing f subject to two equality constraints $g(\vec{x}) = 0, h(\vec{x}) = 0$, we can optimize by solving the equation

$$\nabla f = \lambda \nabla g + \mu \nabla h$$

for some λ, μ. So how do we optimize f when we have multiple constraints, possibly both equalities and inequalities?

Theorem 1 (Karush-Kuhn-Tucker Theorem)**.** *Suppose that f has a local minimizer at $\vec{a}$ satisfying the constraints*

$$\begin{aligned} g_1(\vec{a}) &= k_1 \\ &\vdots \\ g_n(\vec{a}) &= k_n \\ h_1(\vec{a}) &\leq l_1 \\ &\vdots \\ h_m(\vec{a}) &\leq l_m \end{aligned}$$

Let us define the Lagrangian of this problem to be

$$\mathcal{L}(\vec{x}, \vec{y}, \vec{z}) = f(\vec{x}) + \vec{y} \cdot (\vec{g}(\vec{x}) - \vec{k}) + \vec{z} \cdot (\vec{h}(\vec{x}) - \vec{l})$$

If the vectors $\nabla g_1(\vec{a}), \ldots, \nabla g_n(\vec{a}), \nabla h_1(\vec{a}), \ldots, \nabla h_m(\vec{a})$ *are linearly independent, then there exists* $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$ *and* $\vec{\mu} = (\mu_1, \ldots, \mu_m)$ *with each* $\mu_j \geq 0$ *such that* $\vec{a}$ *satisfies the equation*

$$\nabla_{\vec{x}} \mathcal{L}(\vec{a}, \vec{\lambda}, \vec{\mu}) = \nabla f(\vec{a}) + \sum_{i=1}^{n} \lambda_i \nabla g_i(\vec{a}) + \sum_{j=1}^{m} \mu_j \nabla h_j(\vec{a}) = \vec{0}$$

With the additional requirement that for each $1 \leq j \leq m$*,*

$$\mu_j (h_j(\vec{a}) - l_k) = 0$$

Note that the KKT theorem states that it finds a minimizer of f under the constraints. To find a maximizer instead, we can just re-define the function f to be it's negative instead, as the maximizer to f is necessarily the minimizer to $-f$, and vice versa. A similar treatment can be done on constraints of the form $h(\vec{x}) \geq l$.

When illustrating the method of Lagrangian multipliers to optimize f under one equality constraint, an argument via tangent lines/planes/etc. underlies the method. An analogous argument can be made even in the case where we have more than one equality constraint. But how do we treat inequality constraints? Specifically, what is the significance of the condition:

$$\mu(h(\vec{a} - l)) = 0$$

Notice that the above condition holds if and only if at least one of μ or $h(\vec{a}) - l$ holds true; the former implies that the term $\mu \nabla h(\vec{a})$ makes no contribution to $\nabla_{\vec{x}} \mathcal{L}$, while the latter implies that in fact $h(\vec{a}) = l$. Effectively, this is telling us that either $h(\vec{a}) < l$, in which case the constraint $h(\vec{a}) = l$ is not active and we do not need to consider it when identifying (constrained) critical points, or that $h(\vec{a}) = l$ indeed so that we may treat it as an equality constraint. Thus this condition breaks our analysis into two cases: one in which the minimizer lies within the interior of the region determined by $h(\vec{x}) = l$, and the other in which the minimizer lies on the boundary of this region.

Exercises:

1. Maximize $5x + 3y - x^2 - y^2$ subject to the constraints $3x + 2y \leq 9$ and $x + 2y \leq 6$. Notice $\nabla(3x + 2y)$ and $\nabla(x + 2y)$ are never parallel to each other, and thus they are always linearly independent.

2. Maximize $xy + z^2$ subject to the constraints $x + y + z \leq 10$ and $3x + z \leq 24$. Notice that $\nabla(x + y + z)$ and $\nabla(3x + z)$ are never parallel to each other, and thus they are always linearly independent.