Lecture 11, Oct 13, 2023

Equality-Constrained Optimization

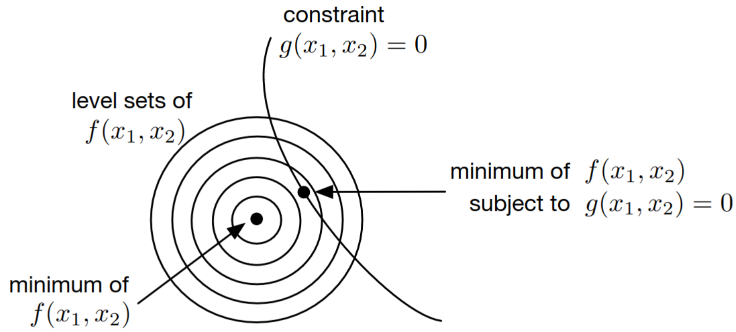

Figure 1: Illustration of equality-constrained optimization.

Definition

A *feasible point* of a constrained optimization problem (minimize $f(\boldsymbol{x})$ subject to $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{h}(\boldsymbol{x}) \geq \boldsymbol{0}$) is any point $\boldsymbol{x}$ satisfying the constraints $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ and $\boldsymbol{h}(\boldsymbol{x}) \geq \boldsymbol{0}$.

The *feasible set* is the set of all feasible points.

Definition

A *critical point* of constrained optimization is a local maximum, minimum, or saddle point of f within the feasible set.

- Consider a simpler problem of minimizing $f(\boldsymbol{x})$ subject to a single equality constraint $g(\boldsymbol{x}) = 0$
  - The feasible set is $S_0 = \{\, \boldsymbol{x} \mid g(\boldsymbol{x}) = 0 \,\}$
  - Consider taking a small step $\Delta\boldsymbol{x}$ from a feasible point; to stay on the level curve $g(\boldsymbol{x}) = 0$, this step must be taken in a directional orthogonal to $\vec{\nabla} g(\boldsymbol{x})$, i.e. $\vec{\nabla} g(\boldsymbol{x}) \Delta\boldsymbol{x} = 0$
  - Now consider taking a small step Δx from a minimum; then we must have $f(\boldsymbol{x}^* + \Delta\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$, if the step is taken such that $g(\boldsymbol{x}) = 0$ is satisfied as above
    - This means $\vec{\nabla} f(\boldsymbol{x}^*) \Delta\boldsymbol{x} \geq 0$ by Taylor expansion
    - But we also have $\vec{\nabla} g(\boldsymbol{x})(-\Delta\boldsymbol{x}) = 0$, so we know that $\vec{\nabla} f(\boldsymbol{x}^*)(-\Delta\boldsymbol{x}) \geq 0$, which means $\vec{\nabla} f(\boldsymbol{x}^*) \Delta\boldsymbol{x} = 0$ is the only way this can be true
- Since $\vec{\nabla} f(x^*) \Delta\boldsymbol{x} = 0 = \vec{\nabla} g(\boldsymbol{x}^*) \Delta\boldsymbol{x}$ we have $\vec{\nabla} f(\boldsymbol{x}^*) = \lambda \vec{\nabla} g(\boldsymbol{x}^*)$ for some λ at a minimum (that is, the gradient of the function and the constraint are parallel)
  - λ is known as the *Lagrange multiplier* (nothing to do with eigenvalues)
- Let the *Lagrangian* be $\Lambda(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda g(\boldsymbol{x})$, then the unconstrained stationary points of $\Lambda(\boldsymbol{x}, \lambda)$ with respect to both λ and $\boldsymbol{x}$ are critical points of the optimization problem, since they satisfy:
  - $\dfrac{\partial \Lambda}{\partial \lambda} = -g(\boldsymbol{x}) = 0$
  - $\dfrac{\partial \Lambda}{\partial \boldsymbol{x}} = \vec{\nabla} f(\boldsymbol{x}) - \lambda \vec{\nabla} g(\boldsymbol{x}) = 0$
  - This allows us to convert a constrained optimization problem to an unconstrained one, so that any unconstrained solver we saw earlier can be used
  - This can be extended to multiple equality constraints

Theorem

Method of Lagrange Multipliers: The critical points of the equality-constrained optimization problem of minimizing $f(\boldsymbol{x})$ subject to $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ are the unconstrained critical points of the *Lagrangian*:

$$\min_{\boldsymbol{x},\boldsymbol{\lambda}} \Lambda(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x})$$

Note that the Lagrangian is minimized with respect to both $\boldsymbol{x}$ and $\boldsymbol{\lambda}$.

- The Lagrange multiplier gives information about how the objective function changes if the constraints change by a small amount
- Example: maximizing a rectangle's area subject to a fixed perimeter of length 1
  - We wish to minimize $f(\boldsymbol{x}) = f(w, l) = -wl$ subject to the constraint that $g(\boldsymbol{x}) = g(w, l) = 2w + 2l - 1 = 0$
  - Use the Lagrangian: $\Lambda(w, l, \lambda) = -wl - \lambda(2w + 2l - 1)$, which we can optimize without constraints
  - $\frac{\partial \Lambda}{\partial w} = -l^* - 2\lambda^*, \frac{\partial \Lambda}{\partial l} = -w^* - 2\lambda^*, \frac{\partial \Lambda}{\partial \lambda} = 1 - 2w^* - 2l^*$
  - In matrix form: $\begin{bmatrix} 0 & -1 & -2 \\ -1 & 0 & -2 \\ 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} w^* \\ l^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
  - We get the solution $w^* = l^* = \frac{1}{4}, \lambda^* = -\frac{1}{8}$, which is a square as expected
- Example: minimize $f(\boldsymbol{x}) = x_1^2 + x_2^2 = \boldsymbol{x}^T\boldsymbol{x}$ subject to $g(\boldsymbol{x}) = x_1 + x_2 - 1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \boldsymbol{x} - 1 = \boldsymbol{a}^T\boldsymbol{x} - 1 = 0$
  - The level curves are circles centered at the origin, and the constraint is a line passing through $(0, 1)$ and $(1, 0)$
  - $\Lambda = f(\boldsymbol{x}) - \lambda g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} - \lambda(\boldsymbol{a}^T\boldsymbol{x} - 1)$
  - $\frac{\partial \Lambda}{\partial \lambda} = \boldsymbol{a}\boldsymbol{x} - 1 = 0 = g(x)$ (we can always just take the constraint; no need to take this derivative)
  - $\frac{\partial \Lambda}{\partial x} = 2\boldsymbol{x}^T - \lambda\boldsymbol{a}^T = 0 \implies x = \frac{\lambda}{2}\boldsymbol{a}$
  - $\boldsymbol{a}^T\boldsymbol{a}\frac{\lambda}{2} = 1 \implies 2\frac{\lambda}{2} = 1 \implies \lambda = 1, \boldsymbol{x} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$
- Example: $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A} = \begin{bmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix}, g(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} - 1 = 0$
  - The level curves are skewed ellipses around the origin; the constraint is a circle of radius 1
  - $\Lambda = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \lambda(\boldsymbol{x}^T\boldsymbol{x} - 1)$
  - $\frac{\partial \Lambda}{\partial \boldsymbol{x}} = 2\boldsymbol{x}^T\boldsymbol{A} - 2\lambda\boldsymbol{x}^T = 0 \implies \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ which is an eigenvalue equation
    - Substituting this back in: $\Lambda = \boldsymbol{x}^T\boldsymbol{x}\lambda - \lambda(\boldsymbol{x}^T\boldsymbol{x} - 1) = \lambda$
  - The eigenvalues are $\lambda = 1, 4$, so we choose $\lambda = 1$ and its eigenvector normalized to unit length (for constraint g)
  - Note that there are actually two solutions here ($\pm$ on the unit eigenvector)