Constrained optimization

Ali Gooya

Reference:

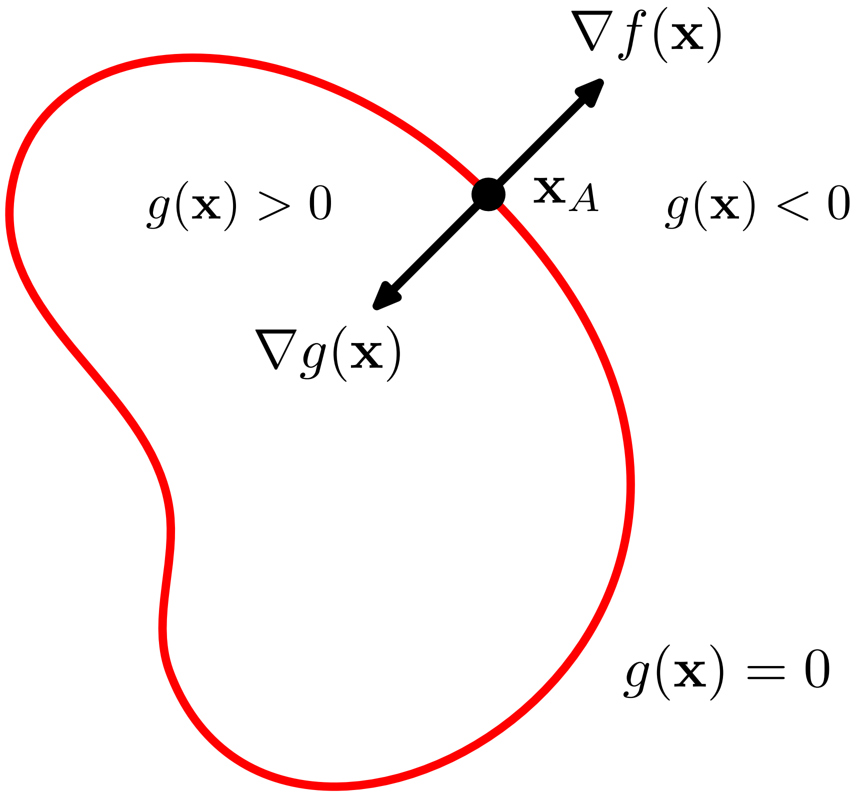
Christopher Bishop, PRML 2006, Appendix D

Equality constraints

- Consider a D dimensional variable $\boldsymbol{x} = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{D})^\mathrm{T}$, we want to maximize $f(\boldsymbol{x})$ with a constraint that $g(\boldsymbol{x}) = 0$.
- We want to show that we need to maximize the Lagrangian given by: $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ with $\lambda \neq 0$
- Note that $\nabla g(\boldsymbol{x})$ is normal to the constraint surface given by $g(\boldsymbol{x}) = 0$
- To see this consider $\boldsymbol{x}_A$ and $\boldsymbol{x}_A + \boldsymbol{\epsilon}$ both placed on the constraint surface.
- For very small $\boldsymbol{\epsilon}$, we can write: $g(\boldsymbol{x}_A + \boldsymbol{\epsilon}) \approx g(\boldsymbol{x}_A) + \boldsymbol{\epsilon}^T \nabla g(\boldsymbol{x}_A)$, but since $g(\boldsymbol{x}_A + \boldsymbol{\epsilon}) = g(\boldsymbol{x}_A) = 0$, then $\boldsymbol{\epsilon}^T \nabla g(\boldsymbol{x}_A) = 0$, which means $\nabla g(\boldsymbol{x})$ is normal to the constraint surface.

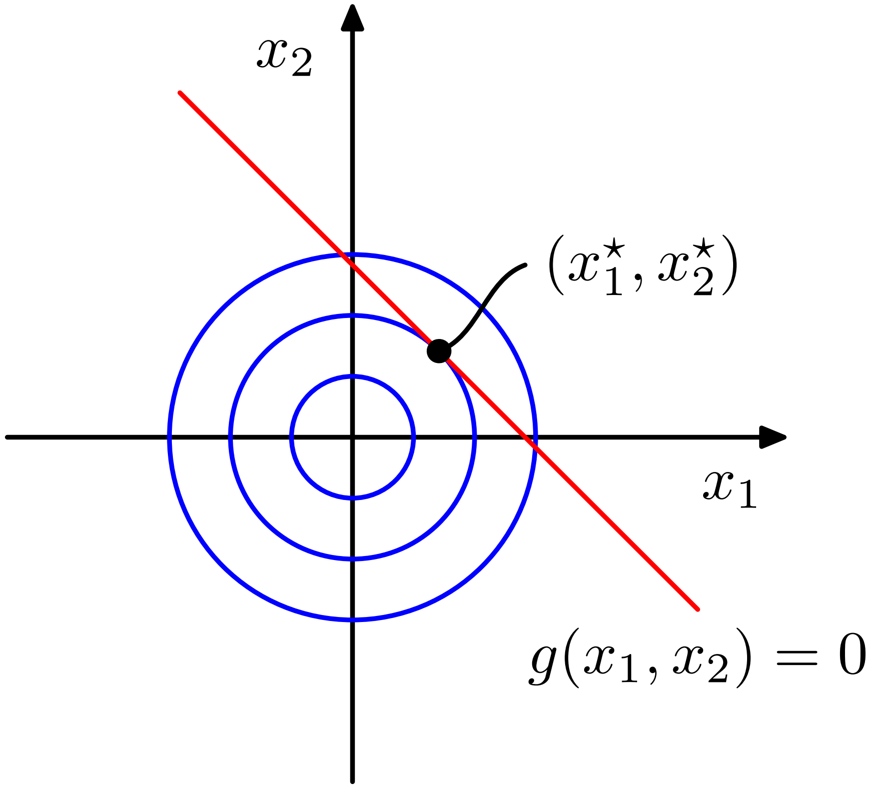
Equality constraints

- Next, we seek a point $\boldsymbol{x}^*$ on the constraint surface such that $f(\boldsymbol{x}^*)$ is maximized.
- Such a point must have the property that the vector $\nabla f(\boldsymbol{x})$ is also orthogonal to the constraint surface, as illustrated below because otherwise we could increase the value of $f(\boldsymbol{x})$ by moving a short distance along the constraint surface. Thus: $\nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = 0$.
- This suggest that we need to consider the gradient of the Lagrangian: $L(\lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ with $\lambda \neq 0$



Example

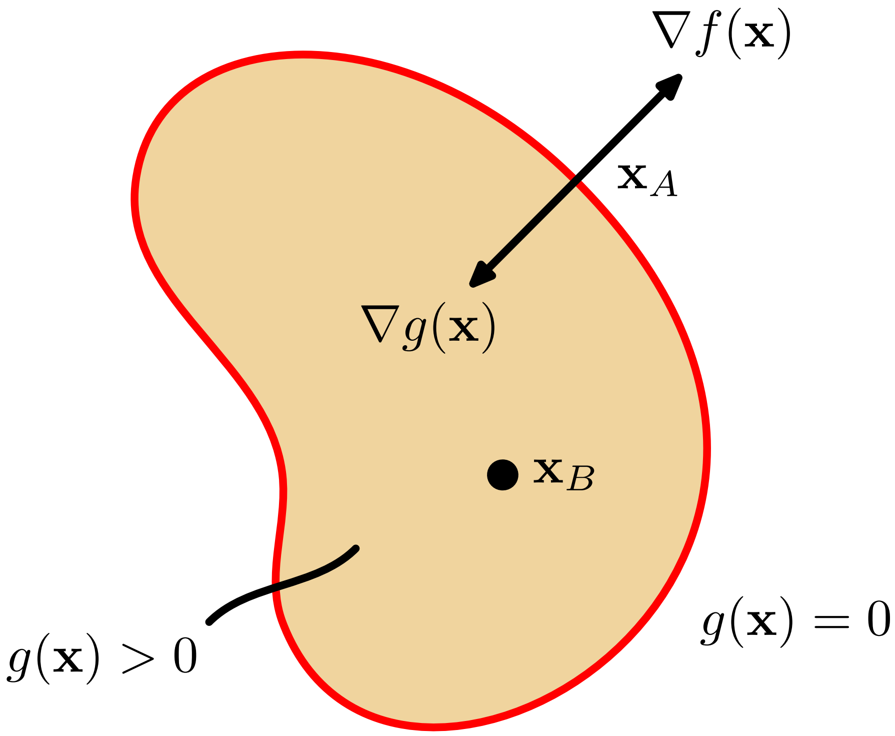
A simple example of the use of Lagrange multipliers in which the aim is to maximize $f(x_1, x_2) = 1 - x_1^2 - x_2^2$ subject to the constraint $g(x_1, x_2) = 0$ where $g(x_1, x_2) = x_1 + x_2 - 1$. The circles show contours of the function $f(x_1, x_2)$, and the diagonal line shows the constraint surface $g(x_1, x_2) = 0$.



$$L(\mathbf{x}, \lambda) = 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1).$$

$$\begin{aligned} -2x_1 + \lambda &= 0 \\ -2x_2 + \lambda &= 0 \\ x_1 + x_2 - 1 &= 0. \end{aligned}$$

Inequality constraints



- Now consider maximising $f(\boldsymbol{x})$ under the constraint $g(\boldsymbol{x}) \geq 0$.
- If stationary point $\boldsymbol{x}_B$ lies in the region where $g(\boldsymbol{x}_B) > 0$, the constraint is ***inactive***, if it lies on the boundary $g(x) = 0$, the constraint is said to be **active**.

- Inactive constraint: $g(\boldsymbol{x})$ plays no role, $\nabla f = 0$ thus, corresponds to stationary point of $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ with $\lambda = 0$
- Active constraint: $g(\boldsymbol{x}) = 0$, similar to equality constraint, $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ with $\lambda \neq 0$. But the sign of λ is important and we should have $\lambda > 0$. Otherwise we can increase f by moving inside region given by $g(\boldsymbol{x}) > 0$ which is contradictory with having active constraint.

Karush-Kuhn-Tucker (KKT) conditions (I)

- Thus, in either case, the solution to the problem of maximizing $f(\boldsymbol{x})$ subject to $g(\boldsymbol{x}) \geq 0$ is obtained by optimizing the Lagrange function $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ and λ subject to the conditions:

$$\begin{aligned} g(\mathbf{x}) &\geqslant 0 \\ \lambda &\geqslant 0 \\ \lambda g(\mathbf{x}) &= 0 \end{aligned}$$

Slackness condition (for $\lambda g(\mathbf{x}) = 0$)

- If we minimize the function $f(\boldsymbol{x})$ subject to $g(\boldsymbol{x}) \leq 0$ then
  - $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$
  - KKT conditions:

$$g(\mathbf{x}) \leq 0$$

$$\lambda \geq 0$$

$$\lambda g(\mathbf{x}) = 0$$

Primal and dual problems

- So to maximize $f(\boldsymbol{x})$, we look for stationary points ($\boldsymbol{x}^*$ and λ^*) of the Lagrangian $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$.
- Maximization of the Lagrangian w.r.t. $\boldsymbol{x}$ is the **primal** problem, whereas the minimization of the reduced Lagrangian, $d(\lambda)$, is the **dual** (always convex) problem.

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}}\, L(\mathbf{x}, \lambda)$$

$$d(\lambda) = L(\mathbf{x}^*, \lambda)$$

$$\lambda^* = \underset{\lambda}{\operatorname{argmin}}\, d(\lambda)$$

Karush-Kuhn-Tucker (KKT) conditions (II)

Finally, it is straightforward to extend the technique of Lagrange multipliers to the case of multiple equality and inequality constraints. Suppose we wish to maximize $f(\mathbf{x})$ subject to $g_j(\mathbf{x}) = 0$ for $j = 1, \ldots, J$, and $h_k(\mathbf{x}) \geqslant 0$ for $k = 1, \ldots, K$. We then introduce Lagrange multipliers $\{\lambda_j\}$ and $\{\mu_k\}$, and then optimize the Lagrangian function given by

$$L(\mathbf{x}, \{\lambda_j\}, \{\mu_k\}) = f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) + \sum_{k=1}^{K} \mu_k h_k(\mathbf{x}) \tag{E.12}$$

subject to $\mu_k \geqslant 0$ and $\mu_k h_k(\mathbf{x}) = 0$ for $k = 1, \ldots, K$. Extensions to constrained functional derivatives are similarly straightforward. For a more detailed discussion of the technique of Lagrange multipliers, see Nocedal and Wright (1999).