ECE 68000: MODERN AUTOMATIC CONTROL

Professor Stan Żak

Minimization subject to equality constraints

Problem statement

- Find a point $\boldsymbol{x} \in \mathbb{R}^N$ that minimizes $f(\boldsymbol{x})$ subject to equality constraints,

$$\left.\begin{array}{rcl} h_1(\boldsymbol{x}) & = & 0 \\ h_2(\boldsymbol{x}) & = & 0 \\ & \vdots & \\ h_M(\boldsymbol{x}) & = & 0 \end{array}\right\}$$

  where $M \leq N$
- Write the above equality constraints in a compact form

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},$$

  where $\boldsymbol{h} : \mathbb{R}^N \to \mathbb{R}^M$
- Surface—the set of points satisfying the above constraints
- The notion of the tangent plane to the surface $S = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ at a point $\boldsymbol{x}^* \in S$

Regular point of the constraints

- A curve on the surface S as a family of points $\boldsymbol{x}(t) \in S$ continuously parameterized by t for $t \in [a, b]$
- The curve is differentiable if $\dot{\boldsymbol{x}}(t) = d\boldsymbol{x}(t)/dt$ exists
- A curve is said to pass through the point $\boldsymbol{x}^* \in S$ if $\boldsymbol{x}^* = \boldsymbol{x}(t^*)$ for some $t^* \in [a, b]$
- The tangent plane to the surface $S = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$ at $\boldsymbol{x}^*$ is the collection of the derivatives at $\boldsymbol{x}^*$ of all differentiable curves on S that pass through $\boldsymbol{x}^*$
- A point $\boldsymbol{x}^*$ satisfying the constraints, that is, $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$, is a regular point of the constraints if the gradient vectors,

$$\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_M(\boldsymbol{x}^*)$$

  are linearly independent

Tangent space

The tangent space at a regular point $\boldsymbol{x}^*$, denoted $T(\boldsymbol{x}^*)$, to the surface $\{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$ at the regular point $\boldsymbol{x}^*$ is

$$T(\boldsymbol{x}^*) = \left\{ \boldsymbol{y} : \begin{bmatrix} \nabla h_1(\boldsymbol{x}^*)^\top \\ \vdots \\ \nabla h_M(\boldsymbol{x}^*)^\top \end{bmatrix} \boldsymbol{y} = \mathbf{0} \right\}$$

The first-order necessary condition (FONC) for function minimization subject to equality constraints

Theorem

Let $\boldsymbol{x}^*$ be a local minimizer (or maximizer) of f subject to the constraints $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$ and let $\boldsymbol{x}^*$ be a regular point of the constraints. Then there exists a vector $\boldsymbol{\lambda}^*$ such that

$$\nabla f\left(\boldsymbol{x}^*\right) + \begin{bmatrix} \nabla h_1\left(\boldsymbol{x}^*\right) & \cdots & \nabla h_M\left(\boldsymbol{x}^*\right) \end{bmatrix} \boldsymbol{\lambda}^* = \mathbf{0}$$

Proof of FONC

- Let $\boldsymbol{x}(t)$ be a differentiable curve passing through $\boldsymbol{x}^*$ on the surface $S = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$ such that $\dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$ where $t^* \in [a, b]$
- Note that $\boldsymbol{y} \in T(\boldsymbol{x}^*)$
- Since $\boldsymbol{x}^*$ is a local minimizer of f on S, we have

$$\left.\frac{d}{dt} f(\boldsymbol{x}(t))\right|_{t=t^*} = 0$$

- Applying the chain rule, gives $\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0$
- Thus $\nabla f(\boldsymbol{x}^*)$ is orthogonal to the tangent space $T(\boldsymbol{x}^*)$
- That is, $\nabla f(\boldsymbol{x}^*)$ is a linear combination of the gradients $\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_M(\boldsymbol{x}^*)$
- This an be expressed as

$$\nabla f(\boldsymbol{x}^*) + \begin{bmatrix} \nabla h_1(\boldsymbol{x}^*) & \cdots & \nabla h_M(\boldsymbol{x}^*) \end{bmatrix} \boldsymbol{\lambda}^* = \mathbf{0}$$

for some constant vector $\boldsymbol{\lambda}^* \in \mathbb{R}^M$ $\square$

The Lagrangian

- The vector $\boldsymbol{\lambda}^*$ is called the vector of Lagrange multipliers
- The Lagrangian associated with the constrained optimization problem,

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x})$$

- Then the FONC can be expressed as

$$\begin{aligned} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \mathbf{0} \end{aligned}$$

- The second of the above condition is equivalent to $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$, that is,

$$\nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$$

- Equivalently the FONC can be written as

$$\begin{aligned} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \mathbf{0} \\ \boldsymbol{h}(\boldsymbol{x}) &= \mathbf{0} \end{aligned}$$

Sequential quadratic programming (SQP)

Apply Newton's method to solve the above system of equations iteratively,

$$\begin{bmatrix} \boldsymbol{x}^{[k+1]} \\ \boldsymbol{\lambda}^{[k+1]} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}^{[k]} \\ \boldsymbol{\lambda}^{[k]} \end{bmatrix} + \begin{bmatrix} \boldsymbol{d}^{[k]} \\ \boldsymbol{y}^{[k]} \end{bmatrix},$$

where $\boldsymbol{d}^{[k]}$ and $\boldsymbol{y}^{[k]}$ are obtained by solving the matrix equation,

$$\begin{bmatrix} \boldsymbol{L}\left(\boldsymbol{x}^{[k]}, \boldsymbol{\lambda}^{[k]}\right) & D\boldsymbol{h}\left(\boldsymbol{x}^{[k]}\right)^\top \\ D\boldsymbol{h}\left(\boldsymbol{x}^{[k]}\right) & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}^{[k]} \\ \boldsymbol{y}^{[k]} \end{bmatrix} = \begin{bmatrix} -\nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{\lambda}) \\ -\boldsymbol{h}\left(\boldsymbol{x}^{[k]}\right) \end{bmatrix},$$

where $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is the Hessian of $l(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$, and $D\boldsymbol{h}(\boldsymbol{x})$ is the Jacobian matrix of $\boldsymbol{h}(\boldsymbol{x})$, that is,

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^\top \\ \vdots \\ \nabla h_M(\boldsymbol{x})^\top \end{bmatrix}$$

The first-order Lagrangian algorithm

- The first-order Lagrangian algorithm for the optimization problem involving minimizing f subject to the equality constraints, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, has the form,

$$\begin{aligned}
\boldsymbol{x}^{[k+1]} &= \boldsymbol{x}^{[k]} - \alpha_k \left(\nabla f\left(\boldsymbol{x}^{[k]}\right) + D\boldsymbol{h}\left(\boldsymbol{x}^{[k]}\right)^\top \boldsymbol{\lambda}^{[k]} \right) \\
\boldsymbol{\lambda}^{[k+1]} &= \boldsymbol{\lambda}^{[k]} + \beta_k \boldsymbol{h}\left(\boldsymbol{x}^{[k]}\right),
\end{aligned}$$

  where α_k and β_k are positive constants
- The update for $\boldsymbol{x}^{[k]}$ is a descent gradient for minimizing the Lagrangian with respect to $\boldsymbol{x}$, while the update for $\boldsymbol{\lambda}^{[k]}$ is a gradient ascent for maximizing the Lagrangian with respect to $\boldsymbol{\lambda}$