Optimality Conditions for Linear Equality Constraints: (11/29)

Let $\mathcal{X}$ denote the feasible set of an optimization problem.

Def: Let $x \in \mathcal{X}$. Then p is a feasible direction if $x + \alpha p \in \mathcal{X}$ for some $\alpha > 0$ small enough.

Def: Let $x^* \in \mathcal{X}$. Then x^* is a constrained local minimum of f if $\partial_p f(x^*) = p^T \nabla f(x^*) \geq 0$ for all feasible directions p.

Let $f \in C^2(\mathbb{R}^n)$ and let $A \in \mathbb{R}^{m \times n}$ be full rank where $m < n$. We want to solve the minimization problem:

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } Ax = b. \end{aligned} \qquad (*)$$

Note: if x is feasible and p is a feasible direction at x, then $p \in \text{null}(A)$.

Pf: Since x is feasible, $Ax = b$. If p is a feasible direction at x, then there exists some $\alpha > 0$ such that $x + \alpha p$ is feasible, i.e. $A(x + \alpha p) = b$. But since $Ax = b$, we have:

$$b = A(x + \alpha p) = Ax + \alpha Ap = b + \alpha Ap,$$

hence $Ap = 0$.

Recall that $\text{null}(A)$ is a vector space of dimension $n - m$ by the rank-nullity theorem. So, to solve (*), we need an initial feasible point and a basis for null(A). Then we simply consider (*) as an unconstrained optimization problem in null(A).

Specifically, let $Z \in \mathbb{R}^{n \times (n-m)}$ span null(A), and let $\bar{x}$ satisfy $A\bar{x} = b$. Then (*) is equivalent to:

$$\underset{v \in \mathbb{R}^{n-m}}{\text{minimize}} \quad f(\bar{x} + Zv), \qquad (**)$$

upon solving (**), we set $x^* = \bar{x} + Zv^*$.

How to compute a basis for the nullspace of A? You can use row-reduction. Another way is to use the QR decomposition.

What we need is a set of vectors $\underbrace{q_{p+1}, \dots, q_n}_{= n-m}$ such that $Aq_i = 0$ for each q_i. If we use Gram-Schmidt to orthogonalize the rows of A to get vectors $q_1, \dots, q_p \in \mathbb{R}^n$, then if we keep running Gram-Schmidt until we get an orthonormal basis for $\mathbb{R}^n$ in the form $q_1, \dots, q_n \in \mathbb{R}^n$, then since:

$$A^T = [q_1 \cdots q_p] R,$$

we will have; for $i = p+1, \dots, n$:

$$A q_i = R^T [q_1 \cdots q_p]^T q_i = R^T \begin{bmatrix} q_1^T q_i \\ \vdots \\ q_p^T q_i \end{bmatrix} = 0.$$

Hence, $q_{m+1}, \dots, q_n$ span null(A).

Let's let $\phi(v) = f(\bar{x} + Zv)$. Then the necessary conditions for optimality of $(**)$ are:

$$\begin{cases} \nabla \phi(v^*) = 0, \\ \nabla^2 \phi(v^*) \text{ is positive semidefinite.} \end{cases}$$

But note:

$$\nabla \phi(v) = Z^T \nabla f(\bar{x} + Zv)$$

$$\nabla^2 \phi(v) = Z^T \nabla^2 f(\bar{x} + Zv) Z.$$

We call $Z^T \nabla f(\bar{x} + Zv)$ the <u>reduced gradient</u> or projected gradient if Z is orthogonal. Likewise, $Z^T \nabla^2 f(\bar{x} + Zv) Z$ is the <u>reduced / projected Hessian</u>. Why do we use this terminology?

Let $Z = [z_1 \cdots z_p]$. Then:

$$Z^T \nabla f = [z_1 \cdots z_p]^T \nabla f = \begin{bmatrix} z_1^T \nabla f \\ \vdots \\ z_p^T \nabla f \end{bmatrix} = \begin{bmatrix} d_{z_1} f \\ \vdots \\ d_{z_p} f \end{bmatrix}.$$

So, $Z^T \nabla f$ gives the gradient of f in the Z-basis, which spans the nullspace. Likewise,

$$Z^T \nabla^2 f Z = \begin{bmatrix} d_{z_1}^2 f & d_{z_1} d_{z_2} f & \cdots & d_{z_1} d_{z_n} f \\ d_{z_2} d_{z_1} f & d_{z_2}^2 f & \cdots & d_{z_2} d_{z_n} f \\ \vdots & \vdots & \ddots & \vdots \\ d_{z_n} d_{z_1} f & d_{z_n} d_{z_2} f & \cdots & d_{z_n}^2 f \end{bmatrix},$$

so $Z^T \nabla f Z$ is the Hessian in the Z-basis.

<u>Necessary conditions</u>: If x^* is a local minimizer for $(*)$ and range(Z) = null(A), then:

$$\begin{cases} Z^T \nabla f(x^*) = 0 \\ Z^T \nabla f(x^*) Z \text{ is positive semidefinite.} \end{cases}$$

Note that these conditions can equivalently be rewritten:

$$d_p f(x^*) = 0 \quad \text{for all } p \in \text{null}(A)$$

$$d_p^2 f(x^*) = p^T \nabla^2 f(x^*) p \geq 0 \quad \text{for all } p \in \text{null}(A)$$

<u>Also note</u>: this does <u>not</u> require $\nabla^2 f(x^*)$ itself to be positive semidefinite.

<u>Sufficient conditions</u>: same as nec. conds. but require $Z^T \nabla^2 f(x^*) Z$ positive <u>definite</u> instead.

Let's observe a connection with Lagrange multipliers now. The Lagrangian for $(*)$ is:

$$L(x, \lambda) = f(x) + (b - Ax)^T \lambda, \quad \lambda \in \mathbb{R}^m.$$

From the KKT conditions, if (x^*, λ^*) are optimal, we have:

$$0 = \left.\frac{\partial L}{\partial x}\right|_{(x^*, \lambda^*)} = \nabla f(x^*) - A^T \lambda^* \Rightarrow \nabla f(x^*) = A^T \lambda^*.$$

On the other hand, we can arrive at this result directly.

If we write $\nabla f(x^*)$ in terms of Z and A^T bases, we have:

$$\nabla f(x^*) = Z\alpha + A^T\beta,$$

for some uniquely determined $\alpha \in \mathbb{R}^{n-m}$, $\beta \in \mathbb{R}^m$. But from the conditions for optimality, we have:

$$0 = Z^T \nabla f(x^*) = Z^T Z \alpha + Z^T A^T \beta.$$

But note that $Z^T Z$ is full rank, and $Z^T A^T = 0$. Hence:

$$Z^T Z \alpha = 0 \Rightarrow \alpha = 0.$$

So we conclude that:

$$\nabla f(x^*) = A^T \beta = A^T \lambda^*.$$

It is also easy to see that $\alpha = v^*$, the unconstrained optimum of $\phi(v) = f(\bar{x} + Zv)$.

This helps give an interpretation of Lagrange multipliers in the case of linear equality constraints.

Another interpretation of Lagrange multipliers

Let's consider $(*)$ again. If we perturb the RHS of the equality constraint by $\delta \in \mathbb{R}^m$, we get:

$$\begin{aligned}&\text{minimize } f(x)\\ &\text{subject to } Ax = b + \delta.\end{aligned}$$

Let (x^*, λ^*) be optimal, and let $\bar{x}$ be close to x^*. Then:

$$f(\bar{x}) = f(x^*) + (\bar{x} - x^*)^T \nabla f(x^*) + O(\|\bar{x} - x^*\|^2)$$

$$= f(x^*) + (\bar{x} - x^*)^T A^T \lambda^* + O(\|\bar{x} - x^*\|^2)$$

$$= f(x^*) + \delta^T \lambda^* + O(\|\bar{x} - x^*\|^2)$$

Since $A(\bar{x} - x^*) = A\bar{x} - b = \delta$.

Hence, if the RHS of the constraints changes by δ, the optimal value of f changes by $\delta^T \lambda^* + O(\|\bar{x} - x^*\|^2)$.