Supplement: Lagrange multipliers.

Recall our setup from class: $f : \mathbb{R}^n \to \mathbb{R}$ is C^1, and $g : \mathbb{R}^n \to \mathbb{R}^m$ is C^1. The function f is your "cost", which you want to maximise (or minimise) subject to the constraint $g = 0$.

The usual strategy is to construct a function

$$H(x, \lambda) = f(x) + \lambda g(x)$$

and observe that critical points of H must satisfy the constraint $g = 0$. Here $\lambda = (\lambda_1 \lambda_2 \dots \lambda_n)$ is a $1 \times n$ matrix (n being the number of coordinates g has, which is the same as the "number of constraints").

The typical situation (see your homework) will leave you with finitely many critical points of H, which you can usually deal with from other considerations. (There also exists a second derivative test involving the augmentation of the Hessian, which is described in your book and required for the last problem.)

Unfortunately, the observation that critical points of H satisfy the constraint g does not rule out the possibility of the existence of a constrained maximum of f which is not a critical point of H! (You saw an example of this in the section, and there is one on your homework.) However, if 0 is a regular value of g, then any constrained maximum of f must in fact correspond to a critical point of H. I proved this for $n = 2$, $m = 1$ in class. But the underlying idea is the same in higher dimensions, provided you know the *Rank Nullity* theorem. Since there are enough other problems to put on your homework this week, I write it up below.

Proposition 1. *Suppose* 0 *is a regular value of* g, *and a constrained (local) maximum of* f *given* $g = 0$ *is attained at the point* x_0. *Then there exists an* $1 \times m$ *matrix* λ_0 *such that*

$$Df_{x_0} + \lambda_0 Dg_{x_0} = 0.$$

Consequently (x_0, λ_0) *is a critical point of* H.

Proof. Let $x = (y, z)$, where $y \in \mathbb{R}^{n-m}$ and $z \in \mathbb{R}^m$. Without loss of generality, we assume that the last m columns of the matrix Dg_{x_0} are linearly independent. Thus, by the implicit function theorem there exists ε, δ and a C^1 function $h : \mathbb{R}^{n-m} \to \mathbb{R}^m$ such that

$$\{g = 0\} \cap B(x_0, \varepsilon) = \{(y, h(z)) \bigm| y \in B(y_0, \varepsilon)\}.$$

Now, f attains a constrained (local) maximum at x_0 if and only if the function $\varphi : \mathbb{R}^{n-m} \to \mathbb{R}^m$ defined by $\varphi(y) = f(y, h(y))$ has a critical point at y_0. Thus

$$0 = D\varphi_{y_0} = Df_{x_0} \begin{pmatrix} I \\ Dh_{y_0} \end{pmatrix}.$$

Equivalently, this means the vector ∇f is in the kernel of the matrix M defined by

$$M \overset{\text{def}}{=} \begin{pmatrix} I & (Dh_{y_0})^* \end{pmatrix}.$$

On the other hand, differentiating the identity $g(y, h(y)) = 0$, we see

$$Dg_{x_0} \begin{pmatrix} I \\ Dh_{y_0} \end{pmatrix} = 0.$$

Equivalently, each of the vectors $\nabla g_1(x_0)$, ..., $\nabla g_m(x_0)$ are in the kernel of the matrix M.

Now let's count dimensions. Note that M is an $(n-m) \times n$ matrix, and the first $n - m$ columns are the same as that of the $(n-m) \times (n-m)$ identity matrix. This forces $\operatorname{rank}(M) = n - m$. So by the Rank-Nullity theorem, $\dim(\ker(M)) = n - \operatorname{rank}(M) = m$.

Observe that the m vectors $\nabla g_1(x_0)$, ..., $\nabla g_m(x_0)$ are all linearly independent (because $\operatorname{rank} Dg_{x_0} = m$). Since they're all in the kernel of M, a subspace of dimension exactly m, these vectors must form a basis of $\ker(M)$. Thus ∇f_{x_0} (which know is in $\ker(M)$) must be a linear combination of these basis vectors. So $\exists \lambda_1, \dots \lambda_m \in \mathbb{R}$ such that

$$\nabla f_{x_0} = \sum_1^m \lambda_i \nabla g_{x_0}.$$

Taking the transpose finishes the proof. □