21-256: Additional notes on the bordered Hessian

November 1, 2017

This short note is intended to illustrate how to use the bordered Hessian in a constrained optimisation problem through examples. Cutting to the chase, let us recall the statement of the theorem (Theorem 7.3.1 in Walker):

Theorem 1. *Suppose $f, g_1, \ldots, g_m$ are second differentiable functions in the variables $x_1, \ldots, x_n$, and that $(a_1, \ldots, a_n, \lambda_1, \ldots, \lambda_m)$ is a critical point of $\mathcal{L}(x_1, \ldots, x_n, y_1, \ldots, y_m) = f - \sum_{i=1}^m yg$. Suppose further that the set of vectors $\{\nabla g_i(a_1, \ldots, a_n) : 1 \leq i \leq m\}$ are linearly independent. If the **last** $n-m$ principal minors of the bordered Hessian $H(a_1, \ldots, a_n, \lambda_1, \ldots, \lambda_m)$ (the Hessian of $\mathcal{L}$ at the above critical point) is such that the smallest minor has sign $(-1)^{m+1}$ and are alternating in sign, then $(a_1, \ldots, a_n)$ is a local constrained maximum of f subject to the constraints $g_i = 0$.*[1]

Let's work through an example to see how to use this theorem:

Problem 1. *Find the maximizer of the objective function $f(w, x, y, z) = -w^2 + -x^2 + -y^2 + -z^2$ subject to the following constraints:*

$$\begin{aligned} g(w, x, y, z) &= 4w - 3y + z + 15 = 0 \\ h(w, x, y, z) &= -2x - y + z + 5 = 0 \end{aligned}$$

In this problem, we would define the Lagrangian function to be

$$\mathcal{L}(w, x, y, z, \lambda, \mu) = f - \lambda g - \mu h$$

To find the critical points of $\mathcal{L}$, we examine the gradient:

$$\nabla \mathcal{L} = \begin{bmatrix} \mathcal{L}_\lambda \\ \mathcal{L}_\mu \\ \mathcal{L}_w \\ \mathcal{L}_x \\ \mathcal{L}_y \\ \mathcal{L}_z \end{bmatrix} = \begin{bmatrix} -g \\ -h \\ f_w - \lambda g_w - \mu h_w \\ f_x - \lambda g_x - \mu h_x \\ f_y - \lambda g_y - \mu h_y \\ f_z - \lambda g_z - \mu h_z \end{bmatrix} = \begin{bmatrix} -4w + 3y - z - 15 \\ 2x + y - z - 5 \\ -2w - 4\lambda \\ -2x + 2\mu \\ -2y + 3\lambda + \mu \\ -2z - \lambda - \mu \end{bmatrix} = \vec{0}$$

Notice that this is a 6 variable linear system with 6 equations. Solving this by Gaussian elimination, we get a unique solution:

$$\begin{aligned} \lambda &= -1 \\ \mu &= -1 \\ w &= 2 \\ x &= -1 \\ y &= -2 \\ z &= 1 \end{aligned}$$

[1] Here is one possible interpretation of the statement to help with the intuition: the function f starts out defined in the n-dimensional space $\mathbb{R}^n$, and to satisfy each constraint $g_i = 0$ we end up with a domain of one lower dimension (e.g. if we want to satisfy the constraint $g(x, y, z) = x^2 + y^2 + z^2 = 1$, then we go from the entire 3-dimensional space to the "2-dimensional" surface of the sphere of radius 1). Thus to satisfy all m constraints simultaneously we end up with a domain which is intuitively of dimension $n - m$ (to be more precise, the domain is a differentiable manifold of dimension $n - m$ embedded in $\mathbb{R}^n$). However, there are some degenerate cases where one of more constraints are satisfied simply by virtue of satisfying other constraints, which is the case where the ∇g_i are not linearly independent. Otherwise, we are then interested in a quadratic approximation of f on the $n - m$ dimensional tangent "hyperplane" (or the affine subspace) to the domain at the critical point, and to see whether this approximation has a maximum, minimum or neither at this point. As this tangent space has dimension $n - m$, this behaviour is captured by the last $n - m$ minors of the Hessian, and in particular if it behaves like a positive definite matrix on these minors, then the critical point would indeed be a maximizer.

Thus the critical point we have found is at $(a_1, a_2, a_3, a_4) = (2, -1, -2, 1)$. Further, we note that ∇g and ∇h are constant vectors everywhere, and in particularly are not parallel, so $\{\nabla g, \nabla h\}$ is indeed linearly independent. So next let us examine the bordered Hessian:

$$H = \begin{bmatrix} \mathcal{L}_{\lambda\lambda} & \mathcal{L}_{\lambda\mu} & \mathcal{L}_{\lambda w} & \mathcal{L}_{\lambda x} & \mathcal{L}_{\lambda y} & \mathcal{L}_{\lambda z} \\ \mathcal{L}_{\mu\lambda} & \mathcal{L}_{\mu\mu} & \mathcal{L}_{\mu w} & \mathcal{L}_{\mu x} & \mathcal{L}_{\mu y} & \mathcal{L}_{\mu z} \\ \mathcal{L}_{w\lambda} & \mathcal{L}_{w\mu} & \mathcal{L}_{ww} & \mathcal{L}_{wx} & \mathcal{L}_{wy} & \mathcal{L}_{wz} \\ \mathcal{L}_{x\lambda} & \mathcal{L}_{x\mu} & \mathcal{L}_{xw} & \mathcal{L}_{xx} & \mathcal{L}_{xy} & \mathcal{L}_{xz} \\ \mathcal{L}_{y\lambda} & \mathcal{L}_{y\mu} & \mathcal{L}_{yw} & \mathcal{L}_{yx} & \mathcal{L}_{yy} & \mathcal{L}_{yz} \\ \mathcal{L}_{z\lambda} & \mathcal{L}_{z\mu} & \mathcal{L}_{zw} & \mathcal{L}_{zx} & \mathcal{L}_{zy} & \mathcal{L}_{zz} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -4 & 0 & 3 & -1 \\ 0 & 0 & 0 & 2 & 1 & -1 \\ -4 & 0 & -2 & 0 & 0 & 0 \\ 0 & 2 & 0 & -2 & 0 & 0 \\ 3 & 1 & 0 & 0 & -2 & 0 \\ -1 & -1 & 0 & 0 & 0 & -2 \end{bmatrix}$$

Note that in this case, again the bordered Hessian is a constant matrix regardless of where the critical point is. As we wish to check for whether (a_1, a_2, a_3, a_4) is a maximizer or not, according to the theorem we will check the last $n - m$ principal minors of the Hessian matrix, where $n = 4$ is the number of variables and $m = 2$ is the number of constraints i.e. we will check the 5th and 6th principal minors of the bordered Hessian:

$$H_5 = \det \begin{bmatrix} 0 & 0 & -4 & 0 & 3 \\ 0 & 0 & 0 & 2 & 1 \\ -4 & 0 & -2 & 0 & 0 \\ 0 & 2 & 0 & -2 & 0 \\ 3 & 1 & 0 & 0 & -2 \end{bmatrix} = -232 < 0$$

$$H_6 = \det(H) = 560 > 0$$

We note that the smallest minor (in this case H_5) has sign $-1 = (-1)^3 = (-1)^{m+1}$, and the signs of the minors alternate (in this case, H_6 has opposite sign). So we satisfy all the conditions of the theorem, and thus we conclude that the point $(2, -1, -2, 1)$ is a constrained maximum of the objective function.

Question 1. *So what happens if I do not satisfy the sign conditions at the end? Is is a minimizer or a saddle point?*

In general, the only thing we can conclude is that the point is NOT a maximizer.

Question 2. *Then how do I find a minimizer?*

Note that $\vec{a}$ is a minimizer of $f(\vec{x})$ if and only if it is a maximizer of $-f(\vec{x})$. So we can perform the algorithm by instead maximizing $-f$.

Question 3. *You've gotta be kidding me! I did all that work to find the critical point and check the Hessian and now you tell me that I have to redo all that work!?*

Not necessarily: note that when you looked for critical points of $\mathcal{L}$, you found both minimizers and maximizers. If you look through the definition of $\nabla\mathcal{L}$ and H, you might notice that you can quickly change everything to the setup to that for $-f$ by changing a few signs, and moreover you will NOT change the critical points this way (although you will need to do a sign change for all the λ's), so it's not a total lost.

Question 4. *What if my constraints are not of the form $g = 0$?*

Rearrange them so that they are!

Question 5. *This is way too much work. Is there a shortcut through all this?*

It depends on the problem. For example, if you know that the constrained domain is bounded (i.e. does not go towards infinity in any direction), then geometric intuition tells us that it must have an absolute maximum and an absolute minimum. So if you found a number of constrained extrema, the one which gives the smallest output must be the minimizer and the one which gives the greatest output must be the maximizer.

In general, there may be various geometric arguments and calculations that could be made to argue that the critical point must be a maximizer/minimizer, but there is no general argument that is guaranteed to work for every case. That's why we are teaching you this stuff!

Question 6. *Another professor / TA / textbook / reference gives a different statement and/or conditions of the theorem! Who should I trust?*

Most likely its going to be an equivalent formulation of the theorem after some transformations. Don't panic!

Question 7. *How is this different from using Lagrange multipliers?*

Note that the process of finding critical points of $\mathcal{L}$ is precisely the same as using Lagrange multipliers to find constrained extrema. The difference is that looking at the bordered Hessian after that allows us to determine if it is a local constrained maximum or a local constrained minimum, which the method of Lagrange multipliers does not tell us.

Question 8. *Why did this note take so long?*

Sorry! Cooking up that example took some time and I had other work over the weekend. Good luck on the midterm!