Chapter 14

# Appendices

## 14.1 IMPLEMENTING CONSTRAINTS WITH LAGRANGE MULTIPLIERS

Consider the problem of minimizing a function of two variables, say, $E(x, y)$, with respect to $x$ and $y$, subject to the constraint that $C(x, y) = 0$. One way to solve this problem is to first use $C(x, y) = 0$ to write $y$ as a function of $x$ and then substitute this function into $E(x, y)$. The resulting function of a single variable $E(x, y(x))$ can now be minimized by setting $dE/dx = 0$. The constraint equation is used to explicitly reduce the number of independent variables.

One problem with this method is that it is rarely possible to solve $C(x, y)$ explicitly for either $y(x)$ or $x(y)$. The method of Lagrange multipliers provides a method of dealing with the constraints in their implicit form.

The constraint equation $C(x, y) = 0$ defines a curve in $(x, y)$ on which the solution must lie (Figure 14.1). As a point is moved along this curve, the value of $E$ changes, achieving a minimum at $(x_0, y_0)$. This point is not the global minimum of $E$, which lies, in general, off the curve. However, it must be at a point at which $\nabla E$ is perpendicular to the tangent $\hat{t}$ to the curve, or else the point could be moved along the curve to further minimize $E$. The tangent is perpendicular to $\nabla C$, so the condition that $\nabla E \perp \hat{t}$ is equivalent to $\nabla E \propto \nabla C$. Writing the proportionality factor as $-\lambda$, we have $\nabla E + \lambda \nabla C = 0$ or $\nabla (E + \lambda C) = 0$. Thus, the constrained minimization of $E$ subject to $C = 0$ is equivalent to the unconstrained minimization of $\Phi = E + \lambda C$, except that a new variable $\lambda$ is introduced that needs to be determined as part of the problem. The two equations $\nabla \Phi = 0$ and $C = 0$ must be solved simultaneously to yield $(x_0, y_0)$ and $\lambda$.

In $N$-dimensions and with $L < N$ constraint equations, each constraint $C_i = 0$ describes an $N - 1$ dimensional surface. Their intersection is an $N$-$L$ dimensional surface. We treat the $L = N - 1$ case here, where the intersection is a curve (the argument for the other cases being similar). As in the two-dimensional case, the condition that a point $(x_0, y_0)$ is at a minimum of $E$ on this curve is that $\nabla E$ is perpendicular to the tangent $\hat{t}$ to the curve. The curve necessarily lies within all of the intersecting surfaces, and so $\hat{t}$ is perpendicular to any surface normal $\nabla C_i$. Thus, the condition that $\nabla E \perp \hat{t}$ is equivalent to the requirement that $\nabla E$ be

Geophysical Data Analysis: Discrete Inverse Theory. DOI: 10.1016/B978-0-12-397160-9.00014-X

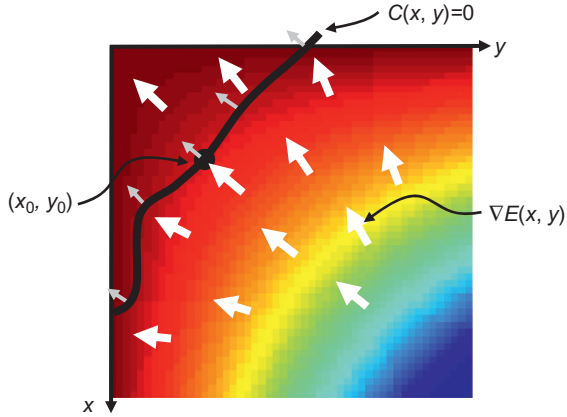


**FIGURE 14.1** Graphical interpretation of the method of Lagrange multipliers, in which the function $E(x, y)$ is minimized subject to the constraint that $C(x, y)=0$. The solution (bold dot) occurs at the point $(x_0, y_0)$ on the curve $C(x, y)=0$, where the perpendicular direction (gray arrows) is parallel to the gradient $\nabla E(x, y)$ (white arrows). At this point, $E$ can only be further minimized by moving the point $(x_0, y_0)$ off of the curve, which is disallowed by the constraint. *MatLab* script gda14_01.

constructed from a linear combination of surface normals, $\nabla C_i$. Calling the coefficients of the linear combination $-\lambda_i$, we have

$$\nabla E = -\sum_{i=1}^{L} \lambda_i \nabla C_i \quad \text{or} \quad \nabla E + \sum_{i=1}^{L} \lambda_i \nabla C_i = 0 \quad \text{or} \quad \nabla \Phi = 0$$
$$\text{with} \quad \Phi = E + \sum_{i=1}^{L} \lambda_i C_i \tag{14.1}$$

As before, the constrained minimization of $E$ subject to $C_i=0$ is equivalent to the unconstrained minimization of $\Phi$, but now $\Phi$ contains one Lagrange multiplier $\lambda_i$ for each of the $L$ constraints.

## 14.2 $L_2$ INVERSE THEORY WITH COMPLEX QUANTITIES

Some inverse problems (especially those that deal with data that have been operated on by Fourier or other transforms) involve complex quantities that are considered random variables. A complex random variable, say $z=x+iy$, has real and imaginary parts, $x$ and $y$, that are themselves real-valued random variables. The probability density function $p(z)$ gives the probability that the real part of $z$ is between $x$ and $x+\Delta x$ and an imaginary part is in between $y$ and $y+\Delta y$. The probability density function is normalized so that its integral over the complex plane is unity:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(z) \mathrm{d}x\mathrm{d}y = 1 \tag{14.2}$$

The mean or expected value of the $z$ is then the obvious generalization of the real-valued case

$$\langle z \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z p(z) \mathrm{d}x\mathrm{d}y \tag{14.3}$$

In general, the real and imaginary parts of $z$ can be of unequal variance and be correlated. However, the uncorrelated, equal variance case, which is called a *circular random variable*, suffices to describe the behavior of many systems. Its variance is defined as

$$\sigma = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (z - \langle z \rangle)^* (z - \langle z \rangle)\, p(z) \mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |z - \langle z \rangle|^2\, p(z) \mathrm{d}x\mathrm{d}y \tag{14.4}$$

Here, the asterisk signifies complex conjugation. A vector $\mathbf{z}$ of $N$ circular random variables, with probability density function $p(\mathbf{z})$, has mean and covariance given by

$$\begin{aligned} \langle \mathbf{z} \rangle_i &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z_i p(\mathbf{z})\, \mathrm{d}^{\mathrm{N}}x\, \mathrm{d}^{\mathrm{N}}y \\ [\mathrm{cov}\mathbf{z}]_{ij} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (z_i - \langle z_i \rangle)^* (z_j - \langle z_j \rangle)\, p(z)\, \mathrm{d}^{\mathrm{N}}x\, \mathrm{d}^{\mathrm{N}}y \end{aligned} \tag{14.5}$$

Note that [cov $\mathbf{z}$] is a Hermetian matrix, that is, a matrix whose transform is its complex conjugate.

The definition of the $L_2$ norm must be changed to accommodate the complex nature of quantities. The appropriate change is to define the squared length of a vector $\mathbf{v}$ to be

$$\|\mathbf{v}\|_2^2 = \mathbf{v}^{\mathrm{H}}\mathbf{v} \tag{14.6}$$

where $\mathbf{v}^{\mathrm{H}}$ is the *Hermetian transpose*, that is, the transpose of the complex conjugate of the vector $\mathbf{v}$. This choice ensures that the norm is a nonnegative real number. When the results of $L_2$ inverse theory are rederived for circular complex vectors, the results are very similar to those derived previously; the only difference is that all the ordinary transposes are replaced by Hermetian transposes. For instance, the least squares solution is

$$\mathbf{m}^{\mathrm{est}} = \left[\mathbf{G}^{\mathrm{H}}\mathbf{G}\right]^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{d} \tag{14.7}$$

Note that all the square symmetric matrices of real inverse theory now become square Hermetian matrices:

$$[\mathrm{cov}\,\mathbf{d}], [\mathrm{cov}\,\mathbf{m}], [\mathbf{G}^{\mathrm{H}}\mathbf{G}], [\mathbf{G}\mathbf{G}^{\mathrm{H}}], \mathrm{etc.} \tag{14.8}$$

Hermetian matrices have real eigenvalues, so no special problems arise when deriving eigenvalues or singular-value decompositions.

The modification made to Householder transformations is also very simple. The requirement that the Hermetian length of a vector be invariant under transformation implies that a unitary transformation must satisfy $\mathbf{T}^{\mathrm{H}}\mathbf{T}=\mathbf{I}$. The most general unitary transformation is therefore $\mathbf{T}=\mathbf{I}-2\mathbf{v}\mathbf{v}^{\mathrm{H}}/\mathbf{v}^{\mathrm{H}}\mathbf{v}$, where $\mathbf{v}$ is any complex vector. The Householder transformation that annihilates the $j$th column of $\mathbf{G}$ below its main diagonal then becomes

$$\mathbf{T}_j = \mathbf{I} - \frac{1}{|\alpha|\left(|\alpha| - |G_{j,j}|\right)} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \left(G_{j,j} - \alpha\right) \\ G_{j+1,j} \\ \vdots \\ G_{N,j} \end{bmatrix} \begin{bmatrix} 0 \cdots 0 & \left(G_{j,j}^* - \alpha^*\right) & G_{j+1,j}^* \cdots G_{N,j}^* \end{bmatrix}$$

$$\text{where } |\alpha| = \sqrt{\sum_{i=j}^{N} |G_{i,j}^2|} \tag{14.9}$$

The phase of $\alpha$ is chosen to be $\pi$ away from the phase of $G_{j,j}$. This choice guarantees that the transformation is in fact a unitary transformation and that the denominator is not zero.