

2.6 Combined Least-Squares and Adjoint

119

2.6.1 Exact Constraints

Consider now a modest generalization of the constrained problem Eq. (2.88) in which the unknowns \mathbf{x} are also meant to satisfy some constraints exactly, or nearly so, for example,

$$\mathbf{Ax} = \mathbf{b}. \tag{2.349} \quad \{35001\}$$

Continued on next page...

In some contexts, (2.349) is referred to as the “model,” a term also employed, confusingly, for the physics defining \mathbf{E} along with the statistics assumed to describe \mathbf{x} , \mathbf{n} . In the end, there is no unique meaning to the term, and only the context is a guide. We will temporarily refer to Eq. (2.349) as “perfect constraints,” as opposed to those involving \mathbf{E} , which generally always have a non-zero noise element.

An example of a model in these terms occurs in acoustic tomography (Chapter 1), where measurements exist of both density and velocity fields, and they are connected by dynamical relations; the errors in the relations are believed to be so much smaller than those in the data, that for practical purposes, the constraints (2.349) might as well be treated as though they are perfect.⁴⁸ But otherwise, the distinction between constraints (2.349) and the observations is an arbitrary one, and the introduction of an error term in the former, no matter how small, removes any particular reason to distinguish them: \mathbf{A} may well be some subset of the rows of \mathbf{E} . What follows can in fact be obtained by imposing the zero noise limit for some of the rows of \mathbf{E} in the solutions already described. Furthermore, whether the model should be satisfied exactly, or should contain a noise element too, is situation dependent. One should be wary of introducing exact equalities into estimation problems, because they carry the strong possibility of introducing small eigenvalues, or near singular relationships, into the solution, and which may dominate the results. Nonetheless, carrying one or more perfect constraints does produce some insight into how the system is behaving.

Several approaches are possible. Consider for example, the objective function,

$$\{35002\} \quad J = (\mathbf{E}\mathbf{x} - \mathbf{y})^T(\mathbf{E}\mathbf{x} - \mathbf{y}) + \gamma^2(\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \quad (2.350)$$

where \mathbf{W} , \mathbf{S} have been previously applied if necessary, and γ^2 is retained as a trade-off parameter. This objective function corresponds to the requirement of a solution of the combined equation sets,

$$\{35003\} \quad \begin{Bmatrix} \mathbf{E} \\ \mathbf{A} \end{Bmatrix} \mathbf{x} + \begin{Bmatrix} \mathbf{n} \\ \mathbf{u} \end{Bmatrix} = \begin{Bmatrix} \mathbf{y} \\ \mathbf{b} \end{Bmatrix} \quad (2.351)$$

in which \mathbf{u} is the model noise, and the weight given to the model is $\gamma^2\mathbf{I}$.) For any finite γ^2 , the perfect constraints are formally “soft” because they are being applied only as a minimized sum of squares. The solution follows immediately from (2.96) with

$$\mathbf{E} \longrightarrow \begin{Bmatrix} \mathbf{E} \\ \alpha\mathbf{A} \end{Bmatrix}, \quad \mathbf{y} \longrightarrow \begin{Bmatrix} \mathbf{y} \\ \alpha\mathbf{b} \end{Bmatrix},$$

assuming the matrix inverse exists. As $\gamma^2 \rightarrow \infty$, the second set of equations is being imposed with arbitrarily great accuracy, and barring numerical issues, becomes as exactly satisfied as one wants (this approach is an example of a “penalty method”).

Alternatively, the model can be imposed as a hard constraint. All prior covariances and scalings having been applied, and Lagrange multipliers introduced, the problem is one with an objective function,

$$J = \mathbf{n}^T \mathbf{n} - 2\boldsymbol{\mu}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{E}\mathbf{x} - \mathbf{y})^T (\mathbf{E}\mathbf{x} - \mathbf{y}) - 2\boldsymbol{\mu}^T (\mathbf{A}\mathbf{x} - \mathbf{b}), \quad (2.352) \quad \{35004\}$$

which is a variant of (2.148). But now, Eq. (2.349) is to be exactly satisfied, and the observations only approximately so.

Setting the derivatives of J with respect to \mathbf{x} , $\boldsymbol{\mu}$ to zero, gives the normal equations,

$$\mathbf{A}^T \boldsymbol{\mu} = \mathbf{E}^T (\mathbf{E}\mathbf{x} - \mathbf{y}) \quad (2.353)$$

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (2.354)$$

Eq. (2.353) represents the adjoint, or “dual” model, for the adjoint or dual solution $\boldsymbol{\mu}$, and the two equation sets are to be solved simultaneously for \mathbf{x} , $\boldsymbol{\mu}$. They are again $M + N$ equations in $M + N$ unknowns (M of the μ_i , N of the x_i), but need not be full-rank. The first set, sometimes referred to as the “adjoint model,” determines $\boldsymbol{\mu}$ from the *difference between* $\mathbf{E}\mathbf{x}$, and \mathbf{y} . The last set is just the exact constraints.

We can most easily solve two extreme cases in Eqs. (2.353, 2.354)—one in which \mathbf{A} is square, $N \times N$, and of full-rank, and one in which \mathbf{E} has this property. In the first case,

$$\tilde{\mathbf{x}} = \mathbf{A}^{-1} \mathbf{b} \quad (2.355) \quad \{35006\}$$

and,

$$\tilde{\boldsymbol{\mu}} = \mathbf{A}^{-T} (\mathbf{E}^T \mathbf{E} \mathbf{A}^{-1} - \mathbf{E}^T) \mathbf{b}. \quad (2.356) \quad \{35008\}$$

Here, the values of $\tilde{\mathbf{x}}$ are completely determined by the full-rank, perfect constraints and the minimization of the deviation from the observations is passive. The Lagrange multipliers or adjoint solution, however, are useful in providing the sensitivity information, $\partial J / \partial \mathbf{b} = 2\boldsymbol{\mu}$, as already discussed. The uncertainty of this solution is zero because of the full rank perfect model assumption (2.354).

In the second case, from (2.353),

$$\tilde{\mathbf{x}} = (\mathbf{E}^T \mathbf{E})^{-1} [\mathbf{E}^T \mathbf{y} + \mathbf{A}^T \boldsymbol{\mu}] \equiv \tilde{\mathbf{x}}_u + (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{A}^T \boldsymbol{\mu}$$

where $\tilde{\mathbf{x}}_u = (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \mathbf{y}$ is the ordinary, unconstrained least-squares solution. Substituting into (2.354) produces,

$$\tilde{\boldsymbol{\mu}} = [\mathbf{A} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{A}^T]^{-1} (\mathbf{b} - \mathbf{A} \tilde{\mathbf{x}}_u) \quad (2.357) \quad \{35009\}$$

and

$$\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_u + (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{A}^T [\mathbf{A} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{A}^T]^{-1} (\mathbf{b} - \mathbf{A} \tilde{\mathbf{x}}_u), \quad (2.358) \quad \{35010\}$$

assuming \mathbf{A} is full-rank underdetermined. The perfect constraints are underdetermined; their range is being fit perfectly, with its nullspace being employed to reduce the misfit to the data as far as possible. The uncertainty of this solution may be written,⁴⁹

$$\mathbf{P} = D^2(\tilde{\mathbf{x}} - \mathbf{x}) = \sigma^2 \left\{ (\mathbf{E}^T \mathbf{E})^{-1} - (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{A}^T [\mathbf{A} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{A}^T]^{-1} \mathbf{A} (\mathbf{E}^T \mathbf{E})^{-1} \right\}, \quad (2.359)$$

which represents a reduction in the uncertainty of the ordinary least-squares solution (first term on the right) by the information in the perfectly known constraints. The presence of \mathbf{A}^{-1} in these solutions is a manifestation of the warning about the possible introduction of components dependent upon small eigenvalues of \mathbf{A} . If neither $\mathbf{E}^T \mathbf{E}$ nor \mathbf{A} is of full-rank one can use, e.g., the SVD with the above solution; the combined \mathbf{E} , \mathbf{A} may be rank deficient, or just determined.

Example

Consider the least-squares problem of solving

$$x_1 + n_1 = 1$$

$$x_2 + n_2 = 1$$

$$x_1 + x_2 + n_3 = 3$$

with uniform, uncorrelated noise of variance 1 in each of the equations. The least-squares solution is then

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1.3333 & 1.3333 \end{bmatrix}^T$$

with uncertainty

$$\mathbf{P} = \begin{bmatrix} 0.6667 & -0.3333 \\ -0.3333 & 0.6667 \end{bmatrix}.$$

But suppose that it is known or desired that $x_1 - x_2 = 1$. Then (2.358) produces $\tilde{\mathbf{x}} = [1.8333 \quad 0.8333]^T$, $\mu = 0.5$, $J' = 0.8333$, with uncertainty

$$\mathbf{P} = \begin{bmatrix} 0.1667 & 0.1667 \\ 0.1667 & 0.1667 \end{bmatrix}.$$

If the constraint is shifted to $x_1 - x_2 = 1.1$, the new solution is $\tilde{\mathbf{x}} = [1.8833 \quad 0.7833]^T$ and the new objective function is $J' = 0.9383$, consistent with the sensitivity deduced from μ .

A more generally useful case occurs when the errors normally expected to be present in the supposedly exact constraints are explicitly acknowledged. If the exact constraints have errors either in the “forcing,” \mathbf{b} , or in a mis-specification of \mathbf{A} , then we write,

$$\mathbf{Ax} = \mathbf{b} + \mathbf{\Gamma u}, \quad (2.360) \quad \{\text{35012}\}$$

assuming, $\langle \mathbf{u} \rangle = 0$, $\langle \mathbf{uu}^T \rangle = \mathbf{Q}$. $\mathbf{\Gamma}$ is a known coefficient matrix included for generality: If for example the errors were thought to be the same in all equations, we could write $\mathbf{\Gamma} = [1, 1, \dots, 1]^T$, and then \mathbf{u} would be just a scalar. Let the dimension of \mathbf{u} be $P \times 1$. Such representations are not unique and more will be said about them in Chapter 4. A hard constraint formulation can still be used, in which (2.360) is still to be exactly satisfied, imposed through an objective function of form,

$$J = (\mathbf{Ex} - \mathbf{y})^T \mathbf{R}_{nn}^{-1} (\mathbf{Ex} - \mathbf{y}) + \mathbf{u}^T \mathbf{Q}^{-1} \mathbf{u} - 2\boldsymbol{\mu}^T (\mathbf{Ax} - \mathbf{b} - \mathbf{\Gamma u}). \quad (2.361) \quad \{\text{35013}\}$$

Here, the noise error covariance matrix has been explicitly included. Finding the normal equations by setting the derivatives with respect to $(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu})$ to zero produces,

$$\mathbf{A}^T \boldsymbol{\mu} = \mathbf{E}^T \mathbf{R}_{nn}^{-1} (\mathbf{Ex} - \mathbf{y}) \quad (2.362)$$

$$\mathbf{\Gamma}^T \boldsymbol{\mu} = \mathbf{Q}^{-1} \mathbf{u} \quad (2.363)$$

$$\mathbf{Ax} + \mathbf{\Gamma u} = \mathbf{b} \quad (2.364)$$

This system is $(2N + P)$ equations in $(2N + P)$ unknowns, where the first equation is again the adjoint system, and dependent upon $\mathbf{Ex} - \mathbf{y}$. Because \mathbf{u} is simple function of the Lagrange multipliers, the system is easily reduced to,

$$\mathbf{A}^T \boldsymbol{\mu} = \mathbf{E}^T \mathbf{R}_{nn}^{-1} (\mathbf{Ex} - \mathbf{y}) \quad (2.365)$$

$$\mathbf{Ax} + \mathbf{\Gamma Q \Gamma}^T \boldsymbol{\mu} = \mathbf{b} \quad (2.366)$$

which is now $2N \times 2N$, the \mathbf{u} having dropped out. If all matrices are full-rank, the solution is immediate; otherwise the SVD is used.

To use a soft constraint methodology, write

$$J = (\mathbf{Ex} - \mathbf{y})^T \mathbf{R}_{nn}^{-1} (\mathbf{Ex} - \mathbf{y}) + (\mathbf{Ax} - \mathbf{b} - \mathbf{\Gamma u})^T \mathbf{Q}^{-1} (\mathbf{Ax} - \mathbf{b} - \mathbf{\Gamma u}), \quad (2.367) \quad \{\text{soft1}\}$$

and find the normal equations. It is again readily confirmed that the solutions using (2.350) or (2.361) are identical, and the hard/soft distinction is seen again to be artificial. The soft constraint method can deal with perfect constraints, by letting $\|\mathbf{Q}^{-1}\| \rightarrow 0$ but stopping when numerical instability sets in. The resulting numerical algorithms fall under the general subject

of “penalty” and “barrier” methods.⁵⁰ Objective functions like ((2.361), 2.367) will be used extensively in Chapter 4.

Example

Consider the partial differential equation

$$\epsilon \nabla^2 \phi + \frac{\partial \phi}{\partial x} = -\sin x \sin y \quad (2.368) \quad \{35014\}$$

A code was written to solve it by finite differences for the case $\epsilon = 0.05$ and $\phi = 0$ on the boundaries $0 \leq x \leq \pi$, $0 \leq y \leq \pi$, as depicted in figure 2.13. The discretized form of the model is then the perfect $N \times N$ constraint system

$$\{35015\} \quad \mathbf{Ax} = \mathbf{b}, \quad \mathbf{x} = \{\phi_{ij}\} \quad (2.369)$$

and \mathbf{b} is the equivalently discretized $-\sin x \sin y$. The theory of partial differential equations shows that this system is full-rank and generally well-behaved. But let us pretend that information is unknown to us, and seek the values \mathbf{x} which makes the objective function

$$\{35016\} \quad J = \mathbf{x}^T \mathbf{x} - 2\boldsymbol{\mu}^T (\mathbf{Ax} - \mathbf{b}) \quad (2.370)$$

stationary with respect to \mathbf{x} , $\boldsymbol{\mu}$, that is the Eqs. (2.353, 2.354) with $\mathbf{E} = \mathbf{I}$, $\mathbf{y} = \mathbf{0}$. Physically, $\mathbf{x}^T \mathbf{x}$ is identified with the solution potential energy. The solution $\boldsymbol{\mu}$, corresponding to the solution of fig. 2.13b is shown in fig. 2.13c. What is the interpretation? The Lagrange multipliers represent the sensitivity of the solution potential energy to perturbations in the forcing field. The sensitivity is greatest in the right-half of the domain, and indeed displays a boundary layer character. A physical interpretation of the Lagrange multipliers can be inferred, given the simple structure of the governing equation (2.368), and the Dirichlet boundary conditions. This equation is not self-adjoint; the adjoint partial differential equation is of form,

$$\{35017c\} \quad \epsilon \nabla^2 \boldsymbol{\nu} - \frac{\partial \boldsymbol{\nu}}{\partial x} = d, \quad (2.371)$$

where d is a forcing term, subject to mixed boundary conditions, and whose discrete form is obtained by taking the transpose of the \mathbf{A} matrix of the discretization (See the Chapter Appendix.) The forward solution exhibits a boundary layer on the left-hand wall, while the adjoint solution has a corresponding behavior in the dual space on the right-hand wall. The structure of the $\boldsymbol{\mu}$ would evidently change if J were changed.⁵¹

The original objective function J is very closely analogous to the Lagrangian (not to be confused with the Lagrange multiplier) in classical mechanics. In mechanics, the gradients of the Lagrangian commonly are virtual forces (forces required to enforce the constraints). The modified Lagrangian, J' , is used in mechanics to impose various physical constraints, and the

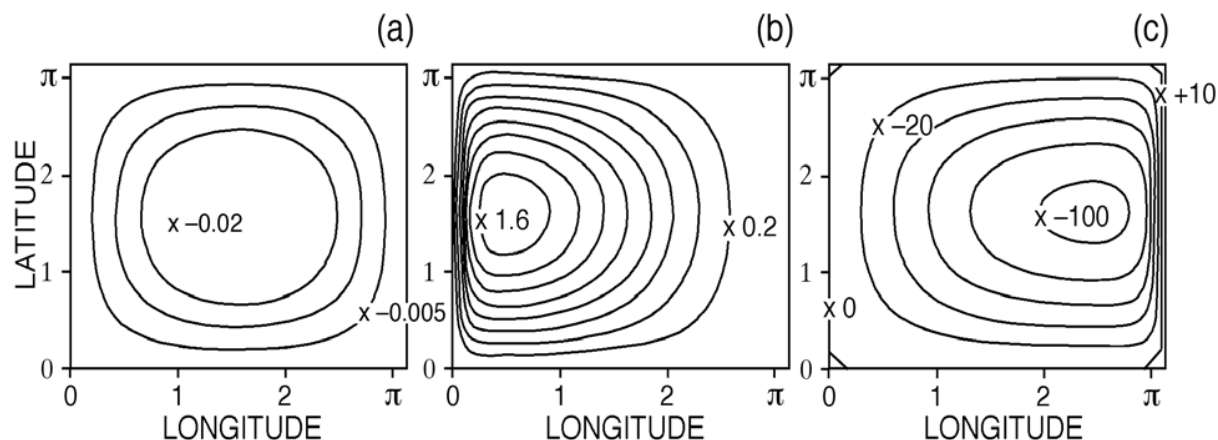


Figure 2.13: Numerical solution of the partial differential equation, Eq. (2.368). Panel (a) shows the imposed symmetric forcing $-\sin x \sin y$. (b) Displays the solution ϕ , and (c) shows the Lagrange multipliers, or adjoint solution, μ and whose structure is a near mirror image of ϕ .

virtual force required to impose the constraints, for example, the demand that a particle follow a particular path, is the Lagrange multiplier.⁵² In an economics/management context, the multipliers are usually called “shadow prices” as they are intimately related to the question of how much profit will change with a shift in the availability or cost of a product ingredient. The terminology “cost function” is a sensible substitute for what we call the “objective function.”

More generally, there is a close connection between the stationarity requirements imposed upon various objective functions throughout this book, and the mathematics of classical mechanics. An elegant Hamiltonian formulation of the material is possible.

2.6.2 Relation to Green Functions⁵³

Consider any linear set of simultaneous equations, involving an arbitrary matrix, \mathbf{A} ,

$$\mathbf{Ax} = \mathbf{b}. \quad (2.372) \quad \{\text{green3}\}$$

Write the adjoint equations for an arbitrary right-hand-side,

$$\mathbf{A}^T \mathbf{z} = \mathbf{r}. \quad (2.373) \quad \{\text{green4}\}$$

Then the simple scalar relation,

$$\mathbf{z}^T \mathbf{Ax} - \mathbf{x}^T \mathbf{A}^T \mathbf{z} = 0 \quad (2.374a) \quad \{\text{bilinear1}\}$$

(the “bilinear identity”) and implies,

$$\mathbf{z}^T \mathbf{b} = \mathbf{x}^T \mathbf{r}. \quad (2.375) \quad \{\text{bilinear2}\}$$

In the special case, $\mathbf{r} = \mathbf{0}$, we have

$$\mathbf{z}^T \mathbf{b} = 0, \quad (2.376)$$

that is, \mathbf{b} , the right-hand side of the original equations (2.372), must be orthogonal to any solution of the homogeneous adjoint equations. (In SVD-terms, this result is nothing but the solvability condition Eq. (2.265).) If \mathbf{A} is of full rank, then there is no non-zero solution to the homogeneous adjoint equations.

Now assume that \mathbf{A} is $N \times N$ of full rank. Add a single equation to (2.372) of the form

$$x_p = \alpha_p \quad (2.377)$$

or

$$\{\text{green5}\} \quad \mathbf{e}_p^T \mathbf{x} = \alpha_p \quad (2.378)$$

where $\mathbf{e}_p = \delta_{ip}$ and α_p is unknown,. We also demand that Eq. (2.372) should remain exactly satisfied. The combined system of (2.372) and (2.378), written as,

$$\mathbf{A}_1 \mathbf{x} = \mathbf{b}_1 \quad (2.379)$$

is overdetermined. If it is to have a solution without any residual, it must still be orthogonal to any solution of the homogeneous adjoint equations,

$$\{\text{green7}\} \quad \mathbf{A}_1^T \mathbf{z} = \mathbf{0}. \quad (2.380)$$

There is only one such solution (because there is only one vector, $\mathbf{z} = \mathbf{u}_{N+1}$, in the null space of \mathbf{A}_1^T). Write $\mathbf{u}_{N+1} = [\mathbf{g}_p, \gamma]^T$, separating out the first N elements of \mathbf{u}_{N+1} , calling them \mathbf{g}_p , and calling the one remaining element γ . Thus Eq. (2.375) is,

$$\mathbf{u}_{N+1}^T \mathbf{b}_1 = \mathbf{g}_p^T \mathbf{b} + \gamma \alpha_p = 0. \quad (2.381)$$

Choose $\gamma = -1$ (any other choice can be absorbed into \mathbf{g}_p). Then,

$$\{\text{green6}\} \quad \alpha_p = \mathbf{g}_p^T \mathbf{b}. \quad (2.382)$$

If \mathbf{g}_p were known, then α_p in (2.382) would be the only value consistent with the solutions to (2.372), and would be the correct value of x_p . But (2.380) is the same as,

$$\{\text{green8}\} \quad \mathbf{A}^T \mathbf{g}_p = \mathbf{e}_p \quad (2.383)$$

(recalling $\gamma = -1$). Because we would like to find *all* elements x_p , we would need to solve (2.383) for all $1 \leq p \leq N$, that is,

$$\{\text{green9}\} \quad \mathbf{A}^T \mathbf{G} = \mathbf{I}_N \quad (2.384)$$

which is N separate problems, each for the corresponding column of $\mathbf{G} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N\}$. Here, \mathbf{G} is the Green function. With \mathbf{G} known, we have immediately,

$$\mathbf{x} = \mathbf{G}^T \mathbf{b}, \quad (2.385)$$

(from Eq. (2.382)). The Green function is an inverse to the adjoint equations (and generalizes in the continuous case to an operator inverse).