Appendix 5

Least-squares Minimization

In this appendix we discuss numerical algorithms for solving linear systems of equations under various constraints. As will be seen such problems are conveniently solved using the SVD.

A5.1 Solution of linear equations

Consider a system of equations of the form $\mathtt{A}\mathbf{x} = \mathbf{b}$. Let $\mathtt{A}$ be an $m \times n$ matrix. There are three possibilities:

(i) If $m < n$ there are more unknowns than equations. In this case, there will not be a unique solution, but rather a vector space of solutions.

(ii) If $m = n$ there will be a unique solution as long as $\mathtt{A}$ is invertible.

(iii) If $m > n$ then there are more equations than unknowns. In general the system will not have a solution unless by chance $\mathbf{b}$ lies in the span of the columns of $\mathtt{A}$.

Least-squares solutions: full-rank case. We consider the case $m \geq n$ and assume for the present that $\mathtt{A}$ is known to be of rank n. If a solution does not exist, then in many cases it still makes sense to seek a vector $\mathbf{x}$ that is closest to providing a solution to the system $\mathtt{A}\mathbf{x} = \mathbf{b}$. In other words, we seek $\mathbf{x}$ such that $\|\mathtt{A}\mathbf{x} - \mathbf{b}\|$ is minimized, where $\|\cdot\|$ represents the vector norm. Such an $\mathbf{x}$ is known as the *least-squares solution* to the over-determined system. The least-squares solution is conveniently found using the SVD as follows.

We seek $\mathbf{x}$ that minimizes $\|\mathtt{A}\mathbf{x} - \mathbf{b}\| = \|\mathtt{UDV}^\mathsf{T}\mathbf{x} - \mathbf{b}\|$. Because of the norm-preserving property of orthogonal transforms, $\|\mathtt{UDV}^\mathsf{T}\mathbf{x} - \mathbf{b}\| = \|\mathtt{DV}^\mathsf{T}\mathbf{x} - \mathtt{U}^\mathsf{T}\mathbf{b}\|$, and this is the quantity that we want to minimize. Writing $\mathbf{y} = \mathtt{V}^\mathsf{T}\mathbf{x}$ and $\mathbf{b}' = \mathtt{U}^\mathsf{T}\mathbf{b}$, the problem becomes one of minimizing $\|\mathtt{D}\mathbf{y} - \mathbf{b}'\|$ where $\mathtt{D}$ is an $m \times n$ matrix with vanishing off-diagonal entries. This set of equations is of the form

$$\left[\begin{array}{cccc} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \\ \hline & & & \\ & & 0 & \\ & & & \end{array}\right] \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_n \\ \hline b'_{n+1} \\ \vdots \\ b'_m \end{pmatrix} .$$

Clearly, the nearest $\mathtt{D}y$ can approach to $\mathbf{b}'$ is the vector $(b'_1, b'_2, \ldots, b'_n, 0, \ldots, 0)^\mathsf{T}$, and this is achieved by setting $y_i = b'_i/d_i$ for $i = 1, \ldots, n$. Note that the assumption $\text{rank}\mathtt{A} = n$ ensures that $d_i \neq 0$. Finally, one retrieves $\mathbf{x}$ from $\mathbf{x} = \mathtt{V}\mathbf{y}$. The complete algorithm is

Objective

Find the least-squares solution to the $m \times n$ set of equations $\mathtt{A}\mathbf{x} = \mathbf{b}$, where $m > n$ and rank $\mathtt{A} = n$.

Algorithm

(i) Find the SVD $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$.
(ii) Set $\mathbf{b}' = \mathtt{U}^\mathsf{T}\mathbf{b}$.
(iii) Find the vector $\mathbf{y}$ defined by $y_i = b'_i/d_i$, where d_i is the i-th diagonal entry of $\mathtt{D}$.
(iv) The solution is $\mathbf{x} = \mathtt{V}\mathbf{y}$.

Algorithm A5.1. *Linear least-squares solution to an over-determined full-rank set of linear equations.*

Deficient-rank systems. Sometimes one is called upon to solve a system of equations that is expected not to be of full column rank. Thus, let $r = \text{rank}\mathtt{A} < n$, where n is the number of columns of $\mathtt{A}$. It is possible that because of noise corruption, the matrix $\mathtt{A}$ actually has rank greater than r, but we wish to enforce the rank r constraint because of theoretical considerations, derived from the particular problem being considered. In this case, there will be an $(n-r)$-parameter family of solutions to the set of equations, where $r = \text{rank}\mathtt{A} < n$. This family of solutions is appropriately solved using the SVD, as follows:

This algorithm gives an $(n-r)$-parameter family (parametrized by the indeterminate values λ_i) of least-squares solutions to the deficient-rank system. The justification of this algorithm is similar to that of algorithm A5.1 for the least-squares solution of full-rank systems.

Systems of unknown rank. In most cases encountered in this book, the rank of a system of linear equations will be known theoretically in advance of solution. If the rank of the system of equations is not known, then one must guess at its rank. In this case, it is appropriate to set singular values that are small compared with the largest

Objective

Find the general solution to a set of equations $\mathtt{A}\mathbf{x} = \mathbf{b}$ where $\mathtt{A}$ is an $m \times n$ matrix of rank $r < n$.

Algorithm

(i) Find the SVD $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$, where the diagonal entries d_i of $\mathtt{D}$ are in descending numerical order.
(ii) Set $\mathbf{b}' = \mathtt{U}^\mathsf{T}\mathbf{b}$.
(iii) Find the vector $\mathbf{y}$ defined by $y_i = b'_i/d_i$ for $i = 1, \ldots, r$, and $y_i = 0$ otherwise.
(iv) The solution $\mathbf{x}$ of minimum norm $\|\mathbf{x}\|$ is $\mathtt{V}\mathbf{y}$.
(v) The general solution is $\mathbf{x} = \mathtt{V}\mathbf{y} + \lambda_{r+1}\mathbf{v}_{r+1} + \ldots + \lambda_n \mathbf{v}_n$, where $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ are the last $n - r$ columns of $\mathtt{V}$.

Algorithm A5.2. *General solution to deficient-rank system*

singular value to zero. Thus, if $d_i/d_0 < \delta$ where δ is a small constant of the order of the machine precision[1] then one sets $y_i = 0$. A least-squares solution is then given by $\mathbf{x} = \mathtt{V}\mathbf{y}$ as before.

A5.2 The pseudo-inverse

Given a square diagonal matrix $\mathtt{D}$, we define its *pseudo-inverse* to be the diagonal matrix $\mathtt{D}^+$ such that

$$D^+_{ii} = \begin{cases} 0 & \text{if } D_{ii} = 0 \\ D_{ii}^{-1} & \text{otherwise.} \end{cases}$$

Now, consider an $m \times n$ matrix $\mathtt{A}$ with $m \geq n$. Let the SVD of $\mathtt{A}$ be $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$. We define the *pseudo-inverse* of $\mathtt{A}$ to be the matrix

$$\mathtt{A}^+ = \mathtt{VD}^+\mathtt{U}^\mathsf{T}. \tag{A5.1}$$

One very simply verifies that the vector $\mathbf{y}$ in algorithm A5.1 or algorithm A5.2 is nothing more than $\mathtt{D}^+\mathbf{b}'$ where $\mathbf{b}' = \mathtt{U}^\mathsf{T}\mathbf{b}$. Thus,

Result A5.1. *The least-squares solution to an $m \times n$ system of equations $\mathtt{A}\mathbf{x} = \mathbf{b}$ of rank n is given by $\mathbf{x} = \mathtt{A}^+\mathbf{b}$. In the case of a deficient-rank system, $\mathbf{x} = \mathtt{A}^+\mathbf{b}$ is the solution that minimizes $\|\mathbf{x}\|$.*

As remarked when discussing the SVD, if $\mathtt{A}$ has fewer rows than columns, then this result may be applied after extending $\mathtt{A}$ to a square matrix by adding rows of zeros.

Symmetric matrices. For symmetric matrices, one may generalize the pseudo-inverse as follows. This generalization was used in chapter 5, section 5.2.3(*p*142) for discussing singular covariance matrices. If $\mathtt{A}$ is a non-invertible symmetric matrix, and $\mathtt{X}^\mathsf{T}\mathtt{AX}$ is invertible, then we write $\mathtt{A}^{+\mathtt{X}} \stackrel{\text{def}}{=} \mathtt{X}(\mathtt{X}^\mathsf{T}\mathtt{AX})^{-1}\mathtt{X}^\mathsf{T}$. One can see that this depends only on the span of the columns of $\mathtt{X}$. In other words, if $\mathtt{X}$ is replaced by $\mathtt{XB}$ for any invertible matrix $\mathtt{B}$, then $\mathtt{A}^{+\mathtt{X}} = \mathtt{A}^{+\mathtt{XB}}$. Otherwise stated, $\mathtt{A}^{+\mathtt{X}}$ depends only on the (left)

[1] Machine precision is the largest floating point value ϵ such that $1.0 + \epsilon = 1.0$.

null-space of $\mathtt{X}$, namely the space of vectors perpendicular to the columns of $\mathtt{X}$. Define the null-space $N_L\mathtt{X} = \{\mathbf{x}^\mathsf{T} \mid \mathbf{x}^\mathsf{T}\mathtt{X} = 0\}$. One finds that under a simple condition, $\mathtt{A}^{+\mathtt{X}}$ is the pseudo-inverse of $\mathtt{A}$:

Result A5.2. *Let $\mathtt{A}$ be a symmetric matrix, then $\mathtt{A}^{+\mathtt{X}} \stackrel{\text{def}}{=} \mathtt{X}(\mathtt{X}^\mathsf{T}\mathtt{A}\mathtt{X})^{-1}\mathtt{X}^\mathsf{T} = \mathtt{A}^+$ if and only if $N_L(\mathtt{X}) = N_L(\mathtt{A})$.*

Only a sketch proof is given. The necessity is obvious, since $N_L(\mathtt{X})$ and $N_L(\mathtt{A})$ are the null-spaces of the left and right sides of the equation. To prove the converse, one may assume that the columns of $\mathtt{X}$ are orthonormal, since as shown above, only the null-space of $\mathtt{X}$ is of importance. Thus, $\mathtt{X}$ may be extended by adding further columns $\mathtt{X}'$ to form an orthogonal matrix $\mathtt{U} = [\mathtt{X}|\mathtt{X}']$. Now, the rows of $\mathtt{X}'^\mathsf{T}$ span the null-space of $\mathtt{X}$, and hence, by assumption, of $\mathtt{A}$. Now, the proof is completed in a few lines by comparing the definition $\mathtt{A}^{+\mathtt{X}} \stackrel{\text{def}}{=} \mathtt{X}(\mathtt{X}^\mathsf{T}\mathtt{A}\mathtt{X})^{-1}\mathtt{X}^\mathsf{T}$ with the definition (A5.1) of the pseudo-inverse.

A5.2.1 Linear least-squares using normal equations

The linear least-squares problem may also be solved by a method involving the so-called *normal equations*. Once more, we consider the set of linear equations $\mathtt{A}\mathbf{x} = \mathbf{b}$ where $\mathtt{A}$ is an $m \times n$ matrix with $m > n$. In general, no solution $\mathbf{x}$ will exist for this set of equations. Consequently, the task is to find the vector $\mathbf{x}$ that minimizes the norm $\|\mathtt{A}\mathbf{x} - \mathbf{b}\|$. As the vector $\mathbf{x}$ varies over all values, the product $\mathtt{A}\mathbf{x}$ varies over the complete column space of $\mathtt{A}$, that is, the subspace of $\mathbb{R}^m$ spanned by the columns of $\mathtt{A}$. The task therefore is to find the closest vector to $\mathbf{b}$ that lies in the column space of $\mathtt{A}$, where closeness is defined in terms of vector norm. Let $\mathbf{x}$ be the solution to this problem; thus $\mathtt{A}\mathbf{x}$ is the closest point to $\mathbf{b}$. In this case, the difference $\mathtt{A}\mathbf{x} - \mathbf{b}$ must be a vector orthogonal to the column space of $\mathtt{A}$. This means, explicitly, that $\mathtt{A}\mathbf{x} - \mathbf{b}$ is perpendicular to each of the columns of $\mathtt{A}$, and hence $\mathtt{A}^\mathsf{T}(\mathtt{A}\mathbf{x} - \mathbf{b}) = \mathbf{0}$. Multiplying out and separating terms gives an equation

$$(\mathtt{A}^\mathsf{T}\mathtt{A})\mathbf{x} = \mathtt{A}^\mathsf{T}\mathbf{b}. \tag{A5.2}$$

This is a square $n \times n$ set of linear equations, called the *normal equations*. This set of equations may be solved to find the least-squares solution to the problem $\mathtt{A}\mathbf{x} = \mathbf{b}$. Even if $\mathtt{A}$ is not of full rank (rank n), this set of equations should have a solution, since $\mathtt{A}^\mathsf{T}\mathbf{b}$ lies in the column space of $\mathtt{A}^\mathsf{T}\mathtt{A}$. In the case where $\mathtt{A}$ has rank n, the matrix $\mathtt{A}^\mathsf{T}\mathtt{A}$ is

Objective

Find $\mathbf{x}$ that minimizes $\|\mathtt{A}\mathbf{x} - \mathbf{b}\|$.

Algorithm

(i) Solve the normal equations $\mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x} = \mathtt{A}^\mathsf{T}\mathbf{b}$.
(ii) If $\mathtt{A}^\mathsf{T}\mathtt{A}$ is invertible, then the solution is $\mathbf{x} = (\mathtt{A}^\mathsf{T}\mathtt{A})^{-1}\mathtt{A}^\mathsf{T}\mathbf{b}$.

Algorithm A5.3. *Linear least-squares using the normal equations.*

invertible, and so $\mathbf{x}$ may be found by $\mathbf{x} = (\mathtt{A}^\mathsf{T}\mathtt{A})^{-1}\mathtt{A}^\mathsf{T}\mathbf{b}$. Since $\mathbf{x} = \mathtt{A}^+\mathbf{b}$, this implies the following result, which is also easily verified directly:

Result A5.3. *If* $\mathtt{A}$ *is an* $m \times n$ *matrix of rank* n*, then* $\mathtt{A}^+ = (\mathtt{A}^\mathsf{T}\mathtt{A})^{-1}\mathtt{A}^\mathsf{T}$.

This result is useful in theoretical analysis, as well as being a computationally simpler method than using the SVD to compute a pseudo-inverse if n is small compared with m (so that computing the inverse of $(\mathtt{A}^\mathsf{T}\mathtt{A})$ is inexpensive compared to computing the SVD of $\mathtt{A}$).

Vector space norms. One sometimes wishes to minimize $\mathtt{A}\mathbf{x} - \mathbf{b}$ with respect to a different norm on the vector space $\mathbb{R}^n$. The usual norm in a vector space $\mathbb{R}^n$ is given in terms of the usual inner product. Thus, for two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^n$ one may define the inner product $\mathbf{a} \cdot \mathbf{b}$ to be $\mathbf{a}^\mathsf{T}\mathbf{b}$. The norm of a vector $\mathbf{a}$ is then $\|\mathbf{a}\| = (\mathbf{a} \cdot \mathbf{a})^{1/2} = (\mathbf{a}^\mathsf{T}\mathbf{a})^{1/2}$. One notes the properties:

(i) The inner product is a symmetric bilinear form on $\mathbb{R}^n$.
(ii) $\|\mathbf{a}\| > 0$ for all non-zero vectors $\mathbf{a} \in \mathbb{R}^n$.

We say that the inner product is a positive-definite symmetric bilinear form. It is possible to define other inner products on a vector space $\mathbb{R}^n$. Let $\mathtt{C}$ be a real symmetric positive-definite matrix, and define a new inner product $\mathtt{C}(\mathbf{a}, \mathbf{b}) = \mathbf{a}^\mathsf{T}\mathtt{C}\mathbf{b}$. The symmetry of the inner product follows from the symmetry of $\mathtt{C}$. A norm may be defined by $\|\mathbf{a}\|_\mathtt{C} = (\mathbf{a}^\mathsf{T}\mathtt{C}\mathbf{a})^{1/2}$, and this is defined and positive-definite, because $\mathtt{C}$ is assumed to be a positive-definite matrix.

Weighted linear least-squares problems. Sometimes one desires to solve a weighted least-squares problem of the form $\mathtt{A}\mathbf{x} - \mathbf{b} = 0$ by minimizing the $\mathtt{C}$-norm $\|\mathtt{A}\mathbf{x} - \mathbf{b}\|_\mathtt{C}$ of the error. Here $\mathtt{C}$ is a positive-definite symmetric matrix defining an inner product and a norm $\|\cdot\|_\mathtt{C}$ on $\mathbb{R}^n$. As before, one can argue that the minimum error vector $\mathtt{A}\mathbf{x} - \mathbf{b}$ must be orthogonal in the inner product defined by $\mathtt{C}$ to the column space of $\mathtt{A}$. This leads to a requirement $\mathtt{A}^\mathsf{T}\mathtt{C}(\mathtt{A}\mathbf{x} - \mathbf{b}) = 0$. Rearranging this one obtains the weighted normal equations:

$$(\mathtt{A}^\mathsf{T}\mathtt{C}\mathtt{A})\mathbf{x} = \mathtt{A}^\mathsf{T}\mathtt{C}\mathbf{b}. \tag{A5.3}$$

The most common weighting will be where $\mathtt{C}$ is a diagonal matrix, corresponding to independent weights in each of the axial directions in $\mathbb{R}^n$. However, general weighting matrices $\mathtt{C}$ may be used also.

A5.3 Least-squares solution of homogeneous equations

Similar to the previous problem is that of solving a set of equations of the form $\mathtt{A}\mathbf{x} = \mathbf{0}$. This problem comes up frequently in reconstruction problems. We consider the case where there are more equations than unknowns – an over-determined set of equations. The obvious solution $\mathbf{x} = \mathbf{0}$ is not of interest – we seek a non-zero solution to the set of equations. Observe that if $\mathbf{x}$ is a solution to this set of equations, then so is $k\mathbf{x}$ for any scalar k. A reasonable constraint would be to seek a solution for which $\|\mathbf{x}\| = 1$.

In general, such a set of equations will not have an exact solution. Suppose $\mathtt{A}$ has dimension $m \times n$ then there is an exact solution if and only if rank$(\mathtt{A}) < n$ – the matrix $\mathtt{A}$ does not have full column rank. In the absence of an exact solution we will normally seek a least-squares solution. The problem may be stated as

- *Find the* $\mathbf{x}$ *that minimizes* $\|\mathtt{A}\mathbf{x}\|$ *subject to* $\|\mathbf{x}\| = 1$.

This problem is solvable as follows. Let $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$. The problem then requires us to minimize $\|\mathtt{UDV}^\mathsf{T}\mathbf{x}\|$. However, $\|\mathtt{UDV}^\mathsf{T}\mathbf{x}\| = \|\mathtt{DV}^\mathsf{T}\mathbf{x}\|$ and $\|\mathbf{x}\| = \|\mathtt{V}^\mathsf{T}\mathbf{x}\|$. Thus, we need to minimize $\|\mathtt{DV}^\mathsf{T}\mathbf{x}\|$ subject to the condition $\|\mathtt{V}^\mathsf{T}\mathbf{x}\| = 1$. We write $\mathbf{y} = \mathtt{V}^\mathsf{T}\mathbf{x}$, and the problem is: minimize $\|\mathtt{D}\mathbf{y}\|$ subject to $\|\mathbf{y}\| = 1$. Now, $\mathtt{D}$ is a diagonal matrix with its diagonal entries in descending order. It follows that the solution to this problem is $\mathbf{y} = (0, 0, \ldots, 0, 1)^\mathsf{T}$ having one non-zero entry, 1 in the last position. Finally $\mathbf{x} = \mathtt{V}\mathbf{y}$ is simply the last column of $\mathtt{V}$. The method is summarized in algorithm A5.4.

<u>Objective</u>

Given a matrix $\mathtt{A}$ with at least as many rows as columns, find $\mathbf{x}$ that minimizes $\|\mathtt{A}\mathbf{x}\|$ subject to $\|\mathbf{x}\| = 1$.

<u>Solution</u>

$\mathbf{x}$ is the last column of $\mathtt{V}$, where $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$ is the SVD of $\mathtt{A}$.

Algorithm A5.4. *Least-squares solution of a homogeneous system of linear equations.*

As mentioned in section A4.4 the last column of $\mathtt{V}$ may alternatively be described as the eigenvector of $\mathtt{A}^\mathsf{T}\mathtt{A}$ corresponding to the smallest eigenvalue.

A5.4 Least-squares solution to constrained systems

In the previous section, we considered a method of least-squares solution of equations of the form $\mathtt{A}\mathbf{x} = \mathbf{0}$. Such problems may arise from situations where measurements are made on a set of image features. With exact measurements and an exact imaging model the mathematical model predicts an exact solution to this system. In the case of inexact image measurements, or noise, there will not be an exact solution. In this case, it makes sense to find a least-squares solution.

On other occasions, however, some of the equations represented by rows of the matrix $\mathtt{A}$ are derived from precise mathematical constraints, and should be satisfied exactly. This set of constraints may be described by a matrix equation $\mathtt{C}\mathbf{x} = \mathbf{0}$, which should be satisfied exactly. Others of the equations are derived from image measurements and are subject to noise. This leads to a problem of the following sort:

- *Find the* $\mathbf{x}$ *that minimizes* $\|\mathtt{A}\mathbf{x}\|$ *subject to* $\|\mathbf{x}\| = 1$ *and* $\mathtt{C}\mathbf{x} = \mathbf{0}$.

This problem can be solved in the following manner. The condition that $\mathbf{x}$ satisfies $\mathtt{C}\mathbf{x} = \mathbf{0}$ means that $\mathbf{x}$ lies perpendicular to each of the rows of $\mathtt{C}$. The set of all such $\mathbf{x}$ is a vector space called the orthogonal complement of the row space of $\mathtt{C}$. We wish to find this orthogonal complement.

First, if $\mathtt{C}$ has fewer rows than columns, then extend it to a square matrix by adding rows of zero elements. This has no effect on the set of constraints $\mathtt{C}\mathbf{x} = \mathbf{0}$. Now, let $\mathtt{C} = \mathtt{UDV}^\mathsf{T}$ be the Singular Value Decomposition of $\mathtt{C}$, where $\mathtt{D}$ is a diagonal matrix with r non-zero diagonal entries. In this case, $\mathtt{C}$ has rank r and the row-space of $\mathtt{C}$ is generated by the first r rows of $\mathtt{V}^\mathsf{T}$. The orthogonal complement of the row-space of $\mathtt{C}$ consists of the remaining rows of $\mathtt{V}^\mathsf{T}$. Define $\mathtt{C}^\perp$ to be the matrix $\mathtt{V}$ with the first r columns deleted. Then $\mathtt{C}\mathtt{C}^\perp = 0$, and so the set of vectors $\mathbf{x}$ satisfying $\mathtt{C}\mathbf{x} = \mathbf{0}$ is spanned by the columns of $\mathtt{C}^\perp$ and we may write any such $\mathbf{x}$ as $\mathbf{x} = \mathtt{C}^\perp\mathbf{x}'$ for suitable $\mathbf{x}'$. Since $\mathtt{C}^\perp$ has orthogonal columns, one observes that $\|\mathbf{x}\| = \|\mathtt{C}^\perp\mathbf{x}'\| = \|\mathbf{x}'\|$. The minimization problem now becomes

- *Find the* $\mathbf{x}'$ *that minimizes* $\|\mathtt{A}\mathtt{C}^\perp\mathbf{x}'\|$ *subject to* $\|\mathbf{x}'\| = 1$.

This is simply an instance of the problem discussed in section A5.3, solved by algorithm A5.4. The complete algorithm for solution of the constrained minimization problem is given as algorithm A5.5.

Objective

Given an $m \times n$ matrix $\mathtt{A}$ with $m \geq n$, find the vector $\mathbf{x}$ that minimizes $\|\mathtt{A}\mathbf{x}\|$ subject to $\|\mathbf{x}\| = 1$ and $\mathtt{C}\mathbf{x} = \mathbf{0}$.

Algorithm

(i) If $\mathtt{C}$ has fewer rows than columns, then add zero-filled rows to $\mathtt{C}$ to make it square. Compute the SVD $\mathtt{C} = \mathtt{UDV}^\mathsf{T}$ where diagonal entries of $\mathtt{D}$ are sorted with non-zero ones first. Let $\mathtt{C}^\perp$ be the matrix obtained from $\mathtt{V}$ by deleting the first r columns of $\mathtt{V}$, where r is the number of non-zero entries in $\mathtt{D}$ (the rank of $\mathtt{C}$).

(ii) Find the solution to the minimization problem $\mathtt{A}\mathtt{C}^\perp\mathbf{x}' = \mathbf{0}$ using algorithm A5.4. The solution is given by $\mathbf{x} = \mathtt{C}^\perp\mathbf{x}'$.

Algorithm A5.5. *Algorithm for constrained minimization.*

A5.4.1 More constrained minimization

A further constrained minimization problem arises in the algebraic estimation method used frequently throughout this book – for instance, for computation of the fundamental matrix (section 11.3(*p*282)) or the trifocal tensor (section 16.3(*p*395)).

The problem is:

- *Minimize* $\|\mathtt{A}\mathbf{x}\|$ *subject to* $\|\mathbf{x}\| = 1$ *and* $\mathbf{x} = \mathtt{G}\hat{\mathbf{x}}$ *for a given matrix* $\mathtt{G}$ *and some unknown vector* $\hat{\mathbf{x}}$.

Note that this is very similar to the previous minimization problem of section A5.4, which was reduced to the form of the present problem in which the matrix $\mathtt{G}$ had orthonormal columns. The condition that $\mathbf{x} = \mathtt{G}\hat{\mathbf{x}}$ for some $\hat{\mathbf{x}}$ means nothing more than that $\mathbf{x}$ lies in the span of the columns of $\mathtt{G}$. Thus, to solve the present problem using algorithm A5.5, we need only to replace $\mathtt{G}$ by a matrix with the same column space (i.e. the space spanned by the columns), but with orthonormal columns. If $\mathtt{G} = \mathtt{UDV}^\mathsf{T}$ where $\mathtt{D}$ has r non-zero entries (that is $\mathtt{G}$ has rank r), then let $\mathtt{U}'$ be the matrix consisting of the first r columns of $\mathtt{U}$. Then $\mathtt{G}$ and $\mathtt{U}'$ have the same column space. As in section A5.4 the solution is found by setting $\mathbf{x}'$ to be the unit vector that minimizes $\|\mathtt{AU}'\mathbf{x}'\|$, then setting $\mathbf{x} = \mathtt{U}'\mathbf{x}'$.

If $\hat{\mathbf{x}}$ is also required, then it may be obtained by solving $\mathtt{G}\hat{\mathbf{x}} = \mathbf{x} = \mathtt{U}'\mathbf{x}'$. The solution is expressed in terms of the pseudo-inverse (section A5.2) as $\hat{\mathbf{x}} = \mathtt{G}^{+}\mathbf{x} = \mathtt{G}^{+}\mathtt{U}'\mathbf{x}'$; it may not be unique if $\mathtt{G}$ does not have full column rank. Since $\mathtt{G}^{+} = \mathtt{VD}^{+}\mathtt{U}^{\mathsf{T}}$, we may write $\hat{\mathbf{x}} = \mathtt{VD}^{+}\mathtt{U}^{\mathsf{T}}\mathtt{U}'\mathbf{x}'$, which simplifies to $\hat{\mathbf{x}} = \mathtt{V}'\mathtt{D}'^{-1}\mathbf{x}'$ where $\mathtt{V}'$ consists of the first r columns of $\mathtt{V}'$ and $\mathtt{D}'$ is the upper $r \times r$ block of $\mathtt{D}$.

The complete method is summarized in algorithm A5.6.

Objective

Find the vector $\mathbf{x}$ that minimizes $\|\mathtt{A}\mathbf{x}\|$ subject to the conditions $\|\mathbf{x}\| = 1$ and $\mathbf{x} = \mathtt{G}\hat{\mathbf{x}}$, where $\mathtt{G}$ has rank r.

Algorithm

(i) Compute the SVD $\mathtt{G} = \mathtt{UDV}^{\mathsf{T}}$, where the non-zero values of $\mathtt{D}$ appear first down the diagonal.
(ii) Let $\mathtt{U}'$ be the matrix comprising the first r columns of $\mathtt{U}$.
(iii) Find the unit vector $\mathbf{x}'$ that minimizes $\|\mathtt{AU}'\mathbf{x}'\|$, using algorithm A5.4.
(iv) The required solution is $\mathbf{x} = \mathtt{U}'\mathbf{x}'$.
(v) If desired, one may compute $\hat{\mathbf{x}}$ as $\mathtt{V}'\mathtt{D}'^{-1}\mathbf{x}'$, where $\mathtt{V}'$ consists of the first r columns of $\mathtt{V}$ and $\mathtt{D}'$ is the upper $r \times r$ block of $\mathtt{D}$.

Algorithm A5.6. *Algorithm for constrained minimization, subject to a span-space constraint.*

A5.4.2 Yet another minimization problem

A very similar problem is

- *Minimize* $\|\mathtt{A}\mathbf{x}\|$ *subject to a condition* $\|\mathtt{C}\mathbf{x}\| = 1$.

This problem comes up for instance in the solution to the DLT camera calibration problem (section 7.3(p184)). In general it will be the case that $\text{rank}\mathtt{C} < n$ where n is the dimension of the vector $\mathbf{x}$. Geometrically the problem may be thought of as finding the "lowest" point on a quadratic surface (specified by $\mathbf{x}^{\mathsf{T}}\mathtt{A}^{\mathsf{T}}\mathtt{A}\mathbf{x}$), with the constraint that the point must lie on the (inhomogeneous) "conic" $\mathbf{x}^{\mathsf{T}}\mathtt{C}^{\mathsf{T}}\mathtt{C}\mathbf{x} = 1$.

We start by taking the SVD of the matrix $\mathtt{C}$, obtaining $\mathtt{C} = \mathtt{UDV}^{\mathsf{T}}$. The condition $\|\mathtt{UDV}^{\mathsf{T}}\mathbf{x}\| = 1$ is equivalent to $\|\mathtt{DV}^{\mathsf{T}}\mathbf{x}\| = 1$, and it is not necessary to compute $\mathtt{U}$ explicitly. Then writing $\mathbf{x}' = \mathtt{V}^{\mathsf{T}}\mathbf{x}$, the problem becomes: minimize $\|\mathtt{AV}\mathbf{x}'\|$ subject to the condition $\|\mathtt{D}\mathbf{x}'\| = 1$ Writing $\mathtt{A}' = \mathtt{AV}$, this becomes: minimize $\|\mathtt{A}'\mathbf{x}'\|$ subject to $\|\mathtt{D}\mathbf{x}'\| = 1$. Thus we have reduced to the case where the constraint matrix is a diagonal matrix, $\mathtt{D}$.

We suppose that $\mathtt{D}$ has r non-zero diagonal entries, and s zero entries, where $r + s = n$, the non-zero entries appearing first on the diagonal of $\mathtt{D}$. Then the entries x'_i of $\mathbf{x}'$ for $i > r$ do not affect the value of $\|\mathtt{D}\mathbf{x}'_i\|$, since the corresponding diagonal entries of $\mathtt{D}$ are zero. Then, for a specific choice of the x'_i for $i = 1, \ldots, r$, the other entries x'_i,

$i = r+1, \ldots, n$ should be chosen so as to minimize the value of $\|\mathtt{A}'\mathbf{x}'_i\|$. We write $\mathtt{A}' = [\mathtt{A}'_1 \mid \mathtt{A}'_2]$ where $\mathtt{A}'_1$ consists of the first r columns of $\mathtt{A}'$ and $\mathtt{A}'_2$ consists of the remaining s columns. Similarly, let $\mathbf{x}'_1$ be an r-vector consisting of the first r elements of $\mathbf{x}'$, and let $\mathbf{x}'_2$ consist of the remaining s elements of $\mathbf{x}'$. Further, let $\mathtt{D}_1$ be the $r \times r$ diagonal matrix consisting of the first r diagonal entries of $\mathtt{D}$. Then $\mathtt{A}'\mathbf{x}' = \mathtt{A}'_1\mathbf{x}'_1 + \mathtt{A}'_2\mathbf{x}'_2$, and the minimization problem is to minimize

$$\|\mathtt{A}'_1\mathbf{x}'_1 + \mathtt{A}'_2\mathbf{x}'_2\| \tag{A5.4}$$

subject to the condition $\|\mathtt{D}_1\mathbf{x}'_1\| = 1$. Now, temporarily fixing $\mathbf{x}'_1$, (A5.4) takes the form of a least-squares minimization problem of the type discussed in section A5.1. According to result A5.1, the value of $\mathbf{x}'_2$ that minimizes (A5.4) is $\mathbf{x}'_2 = -\mathtt{A}'^{+}_2\mathtt{A}'_1\mathbf{x}'_1$. Substituting this in (A5.4) gives $\|(\mathtt{A}'_2\mathtt{A}'^{+}_2 - \mathtt{I})\mathtt{A}'_1\mathbf{x}'_1\|$, which we are required to minimize, subject to the condition $\|\mathtt{D}_1\mathbf{x}'_1\| = 1$. Finally, writing $\mathbf{x}'' = \mathtt{D}_1\mathbf{x}'_1$, the problem reduces at last to a problem of the form of the familiar minimization problem of algorithm A5.4.

- *Minimize* $\|(\mathtt{A}'_2\mathtt{A}'^{+}_2 - \mathtt{I})\mathtt{A}'_1\mathtt{D}_1^{-1}\mathbf{x}''_1\|$, *subject to* $\|\mathbf{x}''\| = 1$.

We now summarize the algorithm.

Objective

Minimize $\|\mathtt{A}\mathbf{x}\|$ subject to $\|\mathtt{C}\mathbf{x}\| = 1$.

Algorithm

(i) Compute the SVD $\mathtt{C} = \mathtt{U}\mathtt{D}\mathtt{V}^{\mathsf{T}}$, and write $\mathtt{A}' = \mathtt{A}\mathtt{V}$.
(ii) Suppose rank$\mathtt{D} = r$ and let $\mathtt{A}' = [\mathtt{A}'_1 \mid \mathtt{A}'_2]$ where $\mathtt{A}'_1$ consists of the first r columns of $\mathtt{A}'$, and $\mathtt{A}'_2$ is formed from the remaining columns.
(iii) Let $\mathtt{D}_1$ be the upper $r \times r$ minor of $\mathtt{D}$.
(iv) Compute $\mathtt{A}'' = (\mathtt{A}'_2\mathtt{A}'^{+}_2 - \mathtt{I})\mathtt{A}'_1\mathtt{D}_1^{-1}$. This is an $n \times r$ matrix.
(v) Minimize $\|\mathtt{A}''\mathbf{x}'\|$ subject to $\|\mathbf{x}''\| = 1$ using algorithm A5.4.
(vi) Set $\mathbf{x}'_1 = \mathtt{D}_1^{-1}\mathbf{x}''$, and $\mathbf{x}'_2 = -\mathtt{A}'^{+}_2\mathtt{A}'_1\mathbf{x}'_1$. Let $\mathbf{x}' = \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \end{pmatrix}$.
(vii) The solution is given by $\mathbf{x} = \mathtt{V}\mathbf{x}'$.

Algorithm A5.7. *Least-squares solution of homogeneous equations subject to the constraint* $\|\mathtt{C}\mathbf{x}\| = 1$.