Appendix B

Structure *into Data*

Above, the data was analyzed to find structures, clusters or classes. In this appendix it is assumed that there already exists some knowledge about the system where the data is coming from, and this structural knowledge is applied to enhance the measurements, to get back to the actual behaviors beyond the noisy observations.

There are different kinds of structures that can be utilized: First, there is the *physical structure,* including not only the actual system structure but also the hierarchic structure determined by the instrumentation of the measurement devices; second, there is the *mathematical structure* as determined by theoretical dependencies among variables; and, third, the observed *a posteriori structure* among the measurements themselves can be utilized. Each of these alternatives is illustrated separately in what follows — what is possible and what is not is very much dependent of the practical system being modeled. Here, only a glimpse into these issues can be given, introducing the challenges and possibilities.

The data is manipulated so that the structural constraints are automatically taken care of when models are constructed for that data. Exploitation of the structures typically introduces new constraints among variables; these constraints become visible in the degrees of freedom in the data. In some cases this reduction in degrees of freedom is reflected as explicit reduction of the data vector dimension. However, if this is not done, explicitly reducing the degrees of dreedom in the data makes the data linearly dependent and collapses the analyses that are based on invertibility of covariance matrices — it turns out that multivariate methods are specially valuable when modeling that data.

B.1 Data reconciliation

The system structure and and measurement variables are, of course, closely linked together. The system structure and its parameters determine what is being measured; on the other hand, these measurement realizations are used to determine the parameters. This report mainly concentrates on the issue of how to utilize the measurements to determine the system structure (or, at

least, its parameters). In this appendix, the known a priori structure is used to determine (or adjust) the measurements: Understanding of the physical system structure is utilized for trying to reconstruct the actual variables beneath the noisy measurement data — what the data values probably *should* have been. This kind of interference in the actual measurements is called *data reconciliation.*

Assume that vector ν represents the noisy measurements, and v is the vector of polished variable values after the structural constraints have been taken into account. The problem of finding variable values v near to measurements, subject to a set of (linear) constraints, can be written in the Lagrangian framework as

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2} \cdot (\nu - v)^T R^{-1} (\nu - v) \\ \text{when} & \Gamma v = \gamma. \end{array} \tag{B.1}$$

Here it is assumed that the measurements ν are distributed normally around the (unknown) correct values v having covariance matrix R; minimizing the above criterion gives the maximum likelihood estimates for v (see Chapter 2). The linear constraints are expressed in the form $\Gamma v = \gamma$, where Γ and γ are a matrix and a vector of compatible sizes.

Note that even if the measurements were exactly correct, delivering the momentary variable values with no error at all, data reconciliation can still be motivated: The measurements only give information of temporary nature, they are not necessarily *representative,* they do not necessarily deliver *essential* information. It is the cumulative effect that is relevant; how the quantity has affected the system behavior over the longer sampling interval.

To apply the Lagrangian methodology, one can first construct the Hamiltonian as

$$J(v) = \frac{1}{2} \cdot (\nu - v)^T R^{-1} (\nu - v) + \mu^T \cdot (\Gamma v - \gamma). \tag{B.2}$$

Note that each constraint equation (as determined by individual rows i in Γ and γ) has a multiplier μ_i of its own; above, this set of constraints has been collected into a single matrix expression, μ_i's being collected in vector μ. Minimizing the Hamiltonian gives the following expression for the gradient:

$$\frac{dJ(v)}{dv} = -R^{-1}(\nu - v) + \Gamma^T \mu = 0, \tag{B.3}$$

resulting in

$$v = \nu - R\Gamma^T \mu. \tag{B.4}$$

For eliminating the other unknown μ from the above expression, one needs to utilize the constraint equation $\Gamma v = \gamma$. Recognizing that when (B.3) is multiplied from the left by ΓR, the only term with v can be substituted, resulting in

$$\mu = \left(\Gamma R \Gamma^T\right)^{-1} (\Gamma \nu - \gamma), \tag{B.5}$$

so that one finally has (assuming that $\Gamma R \Gamma^T$ is invertible)

$$v = \nu - R\Gamma^T \left(\Gamma R \Gamma^T\right)^{-1} (\Gamma \nu - \gamma). \tag{B.6}$$

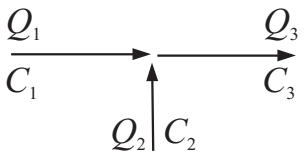

Figure B.1: Three interdependent flows

It is evident that if the measurements ν fulfill the constraints, so that $\Gamma\nu = \gamma$, these values are directly transferred to v, otherwise they are modified according to maximum credibility as expressed in the above formula. Note that the above derivation only modifies data, and constraints directly on the final model parameters cannot be given — see Section B.1.1.

As an example, look at Fig. B.1: Because no accumulation is possible in this subsystem, there must hold

$$Q_1 + Q_2 = Q_3. \tag{B.7}$$

Assuming that the flow values are the only available measurements, so that

$$\nu = \begin{pmatrix} \tilde{Q}_1 & \tilde{Q}_2 & \tilde{Q}_3 \end{pmatrix}^T, \tag{B.8}$$

one has a constraint that can be expressed as

$$\Gamma = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix} \quad \text{with} \quad \gamma = \begin{pmatrix} 0 \end{pmatrix}. \tag{B.9}$$

If there are various measurements of the flows from different time instants, similar constraints have to be written for each time instant separately.

Often the variables are not linearly separable, and the above data reconciliation approaches cannot directly be applied. However, often such problems can still be (approximately) solved. For example, assume that also the concentration values are measured in Fig. B.1, and these values should also be polished. Now the dependencies between variables are highly nonlinear: When the mass balance equations are constructed, in addition to the above volume balance (B.7), one has another weighted average constraint for the solutions:

$$\frac{Q_1 C_1 + Q_2 C_2}{Q_1 + Q_2} = C_3. \tag{B.10}$$

This expression is nonlinear in variables, if all of the measurements are studied simultaneously; on the other hand, if the problem is divided in two separate (suboptimal) optimization tasks, both of these problems are linear. This means that one first solves for the new values for the flow variables as shown above, and when these values are regarded as fixed, one has in the second phase the measurement vector

$$\nu = \begin{pmatrix} \tilde{C}_1 & \tilde{C}_2 & \tilde{C}_3 \end{pmatrix}^T, \tag{B.11}$$

and the mass balance constraint can then be expressed (using the already fixed values for Q_i's) as

$$\Gamma = \begin{pmatrix} \frac{Q_1}{Q_1+Q_2} & \frac{Q_1}{Q_1+Q_2} & -1 \end{pmatrix} \quad \text{with} \quad \gamma = \begin{pmatrix} 0 \end{pmatrix}. \tag{B.12}$$

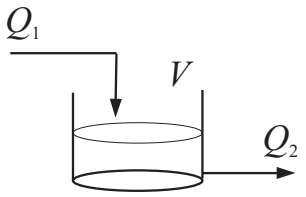

Figure B.2: Dynamics among variables

The final variable vector v can be reconstructed from these data; this approach is not exactly optimal, because the concentration measurements cannot affect the values of the flow variables.

As a more complicated example, study the system in Fig. B.2 that is characterized by the flows Q_1 and Q_2 and the volume V. One knows that the volume at time κ is dependent of the net flow, or, more accurately, the *change* in volume, $V(\kappa+1) - V(\kappa)$, is the same as the effective net flow, $Q_1(\kappa) - Q_2(\kappa)$ multiplied by the sampling interval Δt. If one wants to capture all information that concerns a specific time instant, the data vectors have to be of the form

$$\nu(\kappa) = \begin{pmatrix} \tilde{Q}_1(\kappa) & \tilde{V}(\kappa) & \tilde{Q}_2(\kappa) & \tilde{V}(\kappa+1) \end{pmatrix}^T. \tag{B.13}$$

It is now evident that successive measurement vectors $\nu(\kappa-1)$, $\nu(\kappa)$, $\nu(\kappa+1)$, etc., are linked together because of the shared variables V, and the constraint matrix (consisting of a band of non-zero entries) becomes huge:

$$\left(\begin{array}{ccc|ccc|ccc|c} & & & & & & & & & \\ \cdots & \Delta t & 1 & -\Delta t & & & & & & \cdots \\ & & -1 & & \Delta t & 1 & -\Delta t & & & \\ & & & & & -1 & & & & \end{array}\right) \cdot \left(\begin{array}{c} \vdots \\ \hline Q_1(\kappa) \\ V(\kappa) \\ Q_2(\kappa) \\ \hline Q_1(\kappa+1) \\ V(\kappa+1) \\ Q_2(\kappa+1) \\ \hline \vdots \end{array}\right) = \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}.$$

Note that it is not necessary that all of the quantities needed to construct the structural constraints are measured. The elements in R corresponding to such unmeasured variables can have high values, so that these dummy variables are not weighted. The process can be iterated to reach convergence.

B.1.1 Explicit constraints on parameters

Above, it was the data that was assumed to have some internal structure; it is also possible that the structure exists among the parameters of the final model. For example, assume that the (dynamic single-output) system being modeled can be expressed in the form

$$y_i(k+1) = ay_i(k) + bu(k), \tag{B.14}$$

or

$$y_i(k+1) = \begin{pmatrix} a \\ b \end{pmatrix}^T \begin{pmatrix} y(k) \\ u(k) \end{pmatrix} = F_i^T x(k), \tag{B.15}$$

and, further,

$$Y_i = XF_i. \tag{B.16}$$

Assuming that we *know* that this system represents an ideal mixer, we know that the steady-state gain of the model must equal 1, meaning that there must hold

$$a + b = 1. \tag{B.17}$$

This can be expressed as

$$\begin{pmatrix} 1 & 1 \end{pmatrix} F_i = 1, \tag{B.18}$$

or, more generally, in the form

$$GF_i = g. \tag{B.19}$$

Here, G can also be a matrix and g can be a vector, assuming that there are various constraints to be matched simultaneously. However, G must have more columns than there are rows — otherwise there are no degrees of freedom left for optimization.

To find a model for data, the same procedure as shown above can be applied for constrained optimization:

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\left(Y_i - XF_i\right)^T \left(Y_i - XF_i\right) \\ \text{when} & GF_i = g. \end{array} \tag{B.20}$$

The difference here as compared to the above derivation is that the intended result cannot be assured by data manipulations alone; now the model construction has to be modified. In this sense, the results here differ from other examples in this chapter, but being closely related to data reconciliation, this case is also studied in this context. Differentiating the Hamiltonian

$$\begin{aligned} &\tfrac{d}{dF_i}\left(\tfrac{1}{2}\left(Y_i - XF_i\right)^T \left(Y_i - XF_i\right) - \mu^T\left(GF_i - g\right)\right) \\ &\qquad = X^T XF_i - X^T Y_i - G^T\mu = 0, \end{aligned} \tag{B.21}$$

giving

$$F_i = \left(X^T X\right)^{-1}\left(X^T Y_i + G^T\mu\right). \tag{B.22}$$

To solve for the vector of Lagrange multipliers μ, one can first multiply the above expression from the left by G, and observe that according to the constraint this must equal g:

$$GF_i = G\left(X^T X\right)^{-1} X^T Y_i + G\left(X^T X\right)^{-1} G^T\mu = g, \tag{B.23}$$

so that

$$\mu = \left(G\left(X^T X\right)^{-1} G^T\right)^{-1}\left(g - G\left(X^T X\right)^{-1} X^T Y_i\right), \tag{B.24}$$

and, finally,

$$F_i = \left(X^T X\right)^{-1} \left(X^T Y_i + G^T \left(G \left(X^T X\right)^{-1} G^T \right)^{-1} \left(g - G \left(X^T X\right)^{-1} X^T Y_i \right) \right). \tag{B.25}$$

From the outlook of this expression one can see that the nominal solution of the least-squares minimization is modified by an additive factor that goes to zero if the nominal solution fulfills the given constraint.

What if a more sophisticated regression approach is to be applied, so that the mapping is to go through a subspace spanned by some matrix θ? Assume that the data is first projected onto the latent basis by the mapping matrix $F^1 = \left(\theta^T \theta\right)^{-1} \theta^T$, just as have been done earlier, so that $Z = XF^1$, but the final mapping from the latent basis to the output, F_i^2, is modified from the nominal least-squares fitting so that the overall mapping $F_i = F^1 F_i^2$ fulfills the constraint $GF_i = GF^1 F_i^2 = g$. It is then evident that exactly the above formula (B.25) can be applied if one only selects

$$\begin{array}{lcl} X & \leftarrow & Z = X \left(\theta^T \theta\right)^{-1} \theta^T, \text{ and} \\ G & \leftarrow & GF^1 = G \left(\theta^T \theta\right)^{-1} \theta^T. \end{array} \tag{B.26}$$

The expression (B.25) now actually only gives the mapping F_i^2, so that the final result, or the mapping from input directly to output that fulfills the constraints, with $F^1 = \left(\theta^T \theta\right)^{-1} \theta^T$, is

$$F_i = F^1 \left((XF^1)^T XF^1\right)^{-1} \left((XF^1)^T Y_i + (GF^1)^T \left(GF^1 \left((XF^1)^T XF^1\right)^{-1} (GF^1)^T \right)^{-1} \left(g - GF^1 \left((XF^1)^T XF^1\right)^{-1} (XF^1)^T Y_i \right) \right).$$

B.2 Observing functional hierarchy

Different quantities are measured in different ways, using different kinds of devices. Typically, there is some measurement error present, and one can try to enhance the quality of the data by utilizing various independent (even though somewhat redundant) devices for measuring the same quantity. The number of measurements grows, but — as has been shown — special means are developed to make the models tolerate high dimensionality.

However, the more or less mechanical approaches that were recommended for preprocessing the wealth of measurements do not take into account the functional structure between individual measurements, and the results may be unoptimal. For example, in Chapter **??** it was said that a good approach to reach well-conditioned data is to make the measurements all have the same variance — then the information available from different channels is best balanced. This is a good rule of thumb — but it can be considerably enhanced if additional information is available.

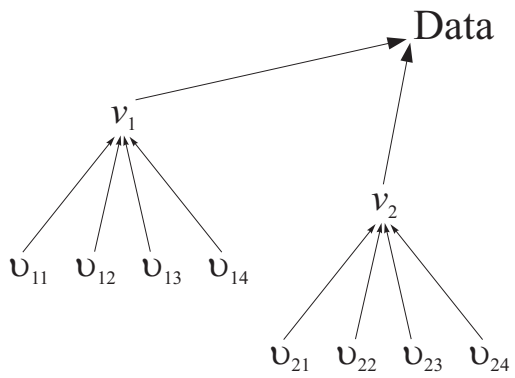

Figure B.3: Hierarchy among measurements

Look at Fig. **??**. It is assumed there that a subset of measurements together try to capture a single physical quantity; the reliability of different measurements (as characterized by the measurement variance) may be different. However, assuming that the devices are independent, all of them deliver some fresh information, and also the less reliable measurements should contribute in the determination of that quantity.

Study an example where n zero-mean measurements of the same quantity v are available, so that $v_i = \nu_i + e_i$ for all $1 \leq i \leq n$. All measurements have characteristic noise properties, so that the variances $\mathrm{var}\{e_i\}$ may vary between measurements; it is assumed that all measurements are unbiased. How should one scale these measurements to reach the best possible estimate $\hat{v} = \sum_{i=1}^{n} w_i v_i$ for v? Scaling of the variables means that the variances are also multiplied, so that $\mathrm{var}\{\hat{v}\} = \sum_i \mathrm{var}\{w_i v_i\} = \sum_i w_i^2 \cdot \mathrm{var}\{v_i\}$. Minimization of the overall variance when the sum of weights is fixed, results in the Lagrangian formulation (see page 20)

$$\left\{ \begin{array}{lcl} f(w_1, \ldots, w_n) & = & \sum_{i=1}^{n} w_i^2 \cdot \mathrm{var}\{v_i\}, \quad \text{and} \\ g(w_1, \ldots, w_n) & = & 1 - \sum_{i=1}^{n} w_i. \end{array} \right. \tag{B.27}$$

The former expression tries to minimize variance, whereas the second expression keeps the estimate $\hat{v}$ unbiased, the sum weights equalling unity. This problem formulation gives

$$J(w_1, \ldots, w_n) = \sum_{i=1}^{n} w_i^2 \cdot \mathrm{var}\{v_i\} + \lambda \cdot \left(1 - \sum_{i=1}^{n} w_i \right), \tag{B.28}$$

so that

$$\left\{ \begin{array}{lcl} \frac{d\, J(w_1, \ldots, w_n)}{d\, w_1} & = & 2 w_1 \cdot \mathrm{var}\{v_1\} - \lambda = 0 \\ & \vdots & \\ \frac{d\, J(w_1, \ldots, w_n)}{d\, w_n} & = & 2 w_n \cdot \mathrm{var}\{v_n\} - \lambda = 0. \end{array} \right. \tag{B.29}$$

Because λ is the same in all of the above equations, it turns out that in the

optimum there must hold for all i

$$w_i \cdot \mathrm{var}\{v_i\} = \textit{constant}. \tag{B.30}$$

This is only possible if the weight is inversely proportional to the corresponding error variance. The optimal weighting between the measurements can also be accomplished as

$$\hat{v} = \alpha \cdot \left(\begin{array}{ccc} \frac{1}{\mathrm{var}\{v_1\}} & \cdots & \frac{1}{\mathrm{var}\{v_n\}} \end{array} \right) \cdot \nu. \tag{B.31}$$

Essentially, forgetting the scalar normalization factor α, the measurements are divided by the observed variances. It seems that division by the *standard deviations* $\sqrt{\mathrm{var}\{v_i\}}$, or normalization of the measurement variances to unity — as proposed later — is not the best way to determine the measurement scaling in this kind of a *sensor fusion* case, where the subset of measurements are tightly coupled together.

Note that still better estimates could be achieved, if not only the variances, but also the *covariances* between measuring devices were taken into account. In such case, the sensor fusion can best be carried out by applying *principal component analysis* (see Chapter 5) for the subset of measurements alone, or, if there is additionally some dynamics in the measurements, by applying *Kalman filter* (see Chapter 9.1). This means that it is not necessary (not even wise) to do all data processing in a centralized manner; if there are clearly independent data analysis subtasks that can be carried out separately, implementing this kind of hierarchical structure in the data processing enhances the overall system robustness and transparency.

B.3 Dimensional analysis

In addition to the above considerations concerning physical structure among the measurements, *mathematical structure* can in some cases also be utilized in a (semi)automatic manner.

Dimensional analysis utilizes the theoretical compatibility properties among variables having different domains: The units have to match to result in mathematically valid expressions. All of the measurements do have some mathematical structure as expressed in terms of basic SI units — distances are written in meters (m), velocities are written as distances divided by time intervals (m/sec), etc. Only such multiplicative combinations of variables are allowed that make all dimensions among additive terms match with each other.

However, to utilize the above idea in practice, one has to make strong assumptions about structure among the variables. It is assumed that the dependency among the n variables can be written in the following form:

$$\nu_1^{f_1} \cdot \cdots \cdot \nu_n^{f_n} = (\text{dimensionless}) \text{ constant}. \tag{B.32}$$

This means that the model structure has to be *multiplicative,* and no additive terms can be allowed in the model (see Sec. **??**). In this case it is only some

distinct combinations of the exponents (parameters) f_i that make the units compatible in (B.32), and this fact is now extensively utilized: Determine a set of parameters so that their all combinations result in valid expressions.

The ideas of dimensional analysis are best explained through an example. Assume that the pressure drop in a tube, Δp, is a function of the tube length l, its diameter d, viscosity of the fluid μ, average speed of the fluid w and its density ρ, so that there exists some function

$$g(l, d, \mu, \rho, w, \Delta p) = 0. \tag{B.33}$$

How to find the functional dependency between the variables based on measurement data? To proceed in the spirit of dimensional analysis, one has to assume that — according to (B.32) — that there holds

$$l^{f_1'} \cdot d^{f_2'} \cdot \mu^{f_3'} \cdot \rho^{f_4'} \cdot w^{f_5'} \cdot \Delta p^{f_6'} = \text{(dimensionless) constant.} \tag{B.34}$$

In principle, even though the functional structure has already been considerably constrained, huge amounts of measurements would still be needed to find all six f_i' parameters (see Sec. **??**). Each variable has the unit of its own, though, and they cannot be freely combined; this reduces the degrees of freedom in the search space. Let us study these units:

Variable	**Unit**
Length l	$[l] = \mathrm{m}$
Diameter d	$[d] = \mathrm{m}$
Viscosity μ	$[\mu] = \mathrm{kg} \cdot \mathrm{m}^{-1} \cdot \mathrm{sec}^{-1}$
Density ρ	$[\rho] = \mathrm{kg} \cdot \mathrm{m}^{-3}$
Velocity w	$[w] = \mathrm{m} \cdot \mathrm{s}^{-1}$
Pressure drop Δp	$[\Delta p] = \mathrm{kg} \cdot \mathrm{m}^{-1} \cdot \mathrm{sec}^{-2}$.

In these expressions, only three basic units are found: *meter* m, *kilogram* kg, and *second* sec. It can be shown (as originally shown in [6]) that if there exist n_0 basic units among the n variables, the degrees of freedom of the formula can be spanned using only $n - n_0$ artificial *dimensionless* variables: The variables increase the degrees of freedom, whereas each basic unit introduces a constraint of its own, reducing the degrees of freedom by one.

Intuitively, the idea is that if all variables that are manipulated are dimensionless, they can be freely multiplied together without problems emerging due to compatibility; any values for exponents are valid (from the mathematical point of view). In this case, one should find those $6 - 3 = 3$ dimensionless variables v_i, so that the same functionality as in (B.33) can be reached in the form

$$g'(v_1, v_2, v_3) = 0, \tag{B.35}$$

or, more specifically, solving (B.34) for v_3, for example,

$$v_3 = f_0 \cdot v_1^{f_1} \cdot v_2^{f_2}. \tag{B.36}$$

When searching for the dimensionless variables v_i, there are various ways to proceed — the methods and also the results are not unique. One practice that

results in easily manageable expressions is to first select $n-n_0$ variables that one thinks are the most relevant, and have the dimensionless variables specifically reflect these (note that the rest of the variables have to contain all base units). For example, if one now wants to find a model for pressure drop, it is reasonable to select Δp among the relevant variables; additionally, let us select l and w. This means that, according to [6], the dimensionless variables are constructed as

$$\begin{cases} v_1 = \rho^{\phi_{11}} \mu^{\phi_{12}} d^{\phi_{13}} \cdot l \\ v_2 = \rho^{\phi_{21}} \mu^{\phi_{22}} d^{\phi_{23}} \cdot w \\ v_3 = \rho^{\phi_{31}} \mu^{\phi_{32}} d^{\phi_{33}} \cdot \Delta p. \end{cases} \tag{B.37}$$

The exponents ϕ_{ij} are now selected so that the units of v_i become dimensionless:

$$\begin{cases} [v_1] = \left(\mathrm{kgm}^{-3}\right)^{\phi_{11}} \left(\mathrm{kgm}^{-1}\mathrm{sec}^{-1}\right)^{\phi_{12}} (\mathrm{m})^{\phi_{13}} \cdot \mathrm{m} = \mathrm{kg}^0\mathrm{m}^0\mathrm{sec}^0 \\ [v_2] = \left(\mathrm{kgm}^{-3}\right)^{\phi_{21}} \left(\mathrm{kgm}^{-1}\mathrm{sec}^{-1}\right)^{\phi_{22}} (\mathrm{m})^{\phi_{23}} \cdot \mathrm{msec}^{-1} = \mathrm{kg}^0\mathrm{m}^0\mathrm{sec}^0 \\ [v_3] = \left(\mathrm{kgm}^{-3}\right)^{\phi_{31}} \left(\mathrm{kgm}^{-1}\mathrm{sec}^{-1}\right)^{\phi_{32}} (\mathrm{m})^{\phi_{33}} \cdot \mathrm{kgm}^{-1}\mathrm{sec}^{-2} = \mathrm{kg}^0\mathrm{m}^0\mathrm{sec}^0, \end{cases}$$

or

$$\begin{cases} (\mathrm{kg})^{\phi_{11}+\phi_{12}} \cdot (\mathrm{m})^{-3\phi_{11}-\phi_{12}+\phi_{13}+1} \cdot (\mathrm{sec})^{-\phi_{12}} = \mathrm{kg}^0\mathrm{m}^0\mathrm{sec}^0 \\ (\mathrm{kg})^{\phi_{21}+\phi_{22}} \cdot (\mathrm{m})^{-3\phi_{21}-\phi_{22}+\phi_{23}+1} \cdot (\mathrm{sec})^{-\phi_{22}-1} = \mathrm{kg}^0\mathrm{m}^0\mathrm{sec}^0 \\ (\mathrm{kg})^{\phi_{31}+\phi_{32}+1} \cdot (\mathrm{m})^{-3\phi_{31}-\phi_{32}+\phi_{33}-1} \cdot (\mathrm{sec})^{-\phi_{32}-2} = \mathrm{kg}^0\mathrm{m}^0\mathrm{sec}^0. \end{cases}$$

From these one can construct a linear set of equations, and the solution for this set becomes

$$\begin{cases} \phi_{11} = 0 \\ \phi_{12} = 0 \\ \phi_{13} = -1, \end{cases} \quad \begin{cases} \phi_{21} = 1 \\ \phi_{22} = -1 \\ \phi_{23} = 1, \end{cases} \quad \text{and} \quad \begin{cases} \phi_{31} = 1 \\ \phi_{32} = -2 \\ \phi_{33} = 2. \end{cases} \tag{B.38}$$

The dimensionless variables also are

$$\begin{cases} v_1 = \rho^0\mu^0 d^{-1} \cdot l = \frac{l}{d} \\ v_2 = \rho^1\mu^{-1} d^1 \cdot w = \frac{wd\rho}{\mu} \\ v_3 = \rho^1\mu^{-2} d^2 \cdot \Delta p = \frac{\Delta p d^2 \rho}{\mu^2}. \end{cases} \tag{B.39}$$

Expression (B.36) can then be written as

$$\left(\frac{d^2\rho\Delta p}{\mu^2}\right) = f_0 \cdot \left(\frac{l}{d}\right)^{f_1'} \cdot \left(\frac{wd\rho}{\mu}\right)^{f_2'}. \tag{B.40}$$

The original problem has been considerably simplified. The expression can be further reduced if there is additional information available: For example, if it is known that the pressure drop is linearly proportional to the tube length, instead of having two separate variables, one can introduce a new independent variable

$$\xi = \Delta p / l \tag{B.41}$$

having the unit $\mathrm{kgm}^{-2}\mathrm{sec}^{-2}$. After this, there only exist $n = 5$ independent variables, and one only needs two dimensionless variables.

It seems that the dimensionless variable v_2 above (accidentally) has the definition of the *Reynold's number* that is familiar from fluid mechanics. This is typical: One often ends up having the same variables when using dimensional analysis — there exist much less freedom among dimensionless variables than there exist dimensioned variables.

It needs to be remembered that this astonishing reduction in the number of variables has its price: First, the assumed functional form (B.34) must be appropriate; second, note that the whole construction collapses if there are, in addition to the variables, some constants that *do have* some dimension. It seems, however, that in fluid mechanics, for example, this kind of assumptions hold, and dimensional analysis is a standard technique in those fields.

Note that in later phases in modeling (as in control engineering in general) the units of the variables are ignored altogether.

B.3.1 Fixing missing data

The above data manipulations were (more or less) well motivated, because the information that was utilized for modifying the data was additional, received from independent external sources — from our *a priori* understanding of the physical or mathematical structure of the system being studied. This last section here, on the other hand, utilizes for modifying the data *a posteriori* structure, determined using a model that has been estimated (as was explained in earlier chapters) by using that *same data.* This means that the steps of data fixing and model construction become an iterative process with some kind of positive feedback[1].

When using the computer, all data structures have to be filled in, there must be no inhomogeneity in the data. In practice, one often has *missing values* among measurements, meaning that some of the variables v_i are unknown; this may be caused, for example, by measurement problems. In the data such problems are often reflected as outliers (see next chapter), lone samples far from the nominal distribution, and it is reasonable to eliminate such erroneous values from the data. however, if there is scarcity with data, or if a contiguous sequence of data is needed for modeling purposes, it may be reasonable to try and fix the incorrect variable values, not to have a "hole" in the data set.

The traditional approach is to substitute the missing values by some kind of average values, either using the average over the whole data set, or calculating the average between the predecessor and the successor (assuming that the same quantity has been measured various times). However, it is clear that such approximations can be extremely crude, and the data distribution may become distorted, resulting in biased models.

The missing value estimate can also be refined iteratively, so that the model

[1]It needs to be kept in mind that fixing data in this way can be extremely dangerous: Using one's intuition about what the data *should* look like, and using such "tailored" data for modeling, makes the model follow this intuition — however incorrect the assumptions were; fresh data should be used where possible!

will be minimally affected by the errors in the fixed variables. One starts with a crude approximation, and step by step makes that value more appropriate, or less conflicting with the other measurements.

As will be shown later, the models that will be constructed later in this report essentially consist of one single matrix F that tries to map a subset of variables v_{in} onto another subset of variables v_{out}, so that there should hold $v_{\text{out}} = F^T v_{\text{in}}$. Assume that the regression model with the mapping matrix $\tilde{F}$ has been constructed using crudely fixed, incorrect data. Then the reconstruction error (the data vectors assumedly containing fixed data in some entries) can be written as

$$e = v_{\text{out}} - \tilde{F}^T v_{\text{in}} = \left(\begin{array}{c|c} I_m & -\tilde{F}^T \end{array} \right) \cdot \left(\frac{v_{\text{out}}}{v_{\text{in}}} \right) = Mv. \tag{B.42}$$

Now, one can rearrange the variables in v so that all variables to be fixed are collected on top, no matter if they belong to the input or output variables. Note that the rows in the matrix M also need to be reordered accordingly. This rearranged set of equations can be written as

$$e = \left(\begin{array}{c|c} M_{\text{NO}} & M_{\text{OK}} \end{array} \right) \cdot \left(\frac{v_{\text{NO}}}{v_{\text{OK}}} \right), \tag{B.43}$$

or

$$e = M_{\text{NO}} v_{\text{NO}} + M_{\text{OK}} v_{\text{OK}}. \tag{B.44}$$

The variables in v_{OK} are assumed to be known *a priori,* whereas the variables in v_{NO} should be modified to better match the model. The next approximation for the missing variables to be fixed can be found when such new values are selected that the matching error e is minimized, so that one has

$$\begin{aligned} \frac{d\left(e^T e\right)}{d\, v_{\text{NO}}} &= \frac{d}{d\, v_{\text{NO}}} \left(M_{\text{NO}} v_{\text{NO}} + M_{\text{OK}} v_{\text{OK}}\right)^T \left(M_{\text{NO}} v_{\text{NO}} + M_{\text{OK}} v_{\text{OK}}\right) \\ &= \frac{d}{d\, v_{\text{NO}}} \left(v_{\text{NO}}^T M_{\text{NO}}^T M_{\text{NO}} v_{\text{NO}} + v_{\text{NO}}^T M_{\text{NO}}^T M_{\text{OK}} v_{\text{OK}} \right. \\ &\qquad \left. + v_{\text{OK}}^T M_{\text{OK}}^T M_{\text{NO}} v_{\text{NO}} + v_{\text{OK}}^T M_{\text{OK}}^T M_{\text{OK}} v_{\text{OK}} \right) \\ &= 2 M_{\text{NO}}^T M_{\text{NO}} v_{\text{NO}} + 2 M_{\text{NO}}^T M_{\text{OK}} v_{\text{OK}} \\ &= \mathbf{0}. \end{aligned}$$

This gives

$$v_{\text{NO}} = -\left(M_{\text{NO}}^T M_{\text{NO}}\right)^\dagger M_{\text{NO}}^T M_{\text{OK}} v_{\text{OK}}. \tag{B.45}$$

So, having found better approximates for the missing values, a new model can be constructed, where the error due to the missing data should be smaller; this refinement procedure can be continued until the parameters converge. The more there are missing values, the more probably the process converges in some local rather than global minimum. Note that the above pseudoinverse may be rank deficient, if too many variables are unknown in a single sample.

Above, it was assumed that one can concentrate on a single issue at a time — now, the data was fixed utilizing the knowledge of some existing structures,

and it is assumed that other issues, like constructing the actual model, can be concentrated on later, separately. The same step-by-step approach has been employed troughout this report, introducing new ideas only after the motivation for them can be understood; this understanding being engineering-like "hands-on" understanding rather than mathematically exhaustive mastering of details. Of course, some level of iteration cannot be avoided: It is clear that in engineering work iteration is necessary — on both conceptual and practical levels — because more can be found in the underlying issues if there is some understanding of the entity. The linear methods that have been elaborated on facilitate fast execution times, so that different kinds of iterative refinement schemes become feasible.

Computer exercises

1. Define original data and constraints as follows:

```
v = [1,1,1]';
R = eye(3);
G = [1,1,1];
g = 1;
```

   Vary the covariance matrix changing the first variance between zero and very high values and study the results:

```
R(1,1) = input('Give variance of the first measurement: ');
RegrReconc(v,R,G,g)
```

2. Test the outlier detection using, for example, the following commands:

```
[X,Y] = dataxy(10,2,2);
X(1,1) = 10;
outl([X,Y]);
```

   Iteratively fix the missing value of the previous exercise by the following commands:

```
Wx = ones(size(X));
Wy = ones(size(Y));
Wx(1,1) = 0;
for i = 1:5
   F = mlr(X,Y);
   [X,Y] = fixval(X,Y,F,Wx,Wy)
end
```

   What may happen in the fixing process if there are too many degrees of freedom, say,

```
[X,Y] = dataxy(10,5,5);
X(1,1) = 10;
```