4 Complex expansions and regularization problems

How can four-wire LV grids be modelled to learn from three-phase voltage data? How can we use pseudo-measurements on single-phase loads to regularize the ill-posed estimation problems that result from the usual lack of real-time metering data in LV grids?

In previous chapters, we have made use of models for both regression and classification that were linear or quadratic in the input features: we have used linear regression approaches to the phase identification and to the state estimation problems; and used a quadratic basis expansion to the loss prediction problem. Both the linear and the quadratic approaches relied upon approximated models that proved accurate and were easy to interpret.

It is not always like that. Sometimes, higher order approximations – higher than the first and second-order Taylor approximations of $h(X)$ – are necessary to obtain good results. Yet, this chapter is not addressing methods for moving into higher order approximations. This is not the source of *complexity* addressed in the context of this chapter.

Complexity in this chapter refers to the order of the relationships between variables, not to the order of the functions that approximate such relationships. The complexity of relationships between network variables in LV grids is quite significant, and therefore the graph structures needed to model LV grids prior knowledge are especially complex. In this chapter we will illustrate how to define graph complex functions h of the inputs X to allow solving two previously addressed problems in new contexts. We readdress the (i) phase identification problem with grid typical output data on unbalanced voltages alone, and (ii) the SE problem in unbalanced three-phase four-wire grids.

The first problem will require the use of a carefully designed function, $h(X)$, to be able to identify connection phases based on bus voltage readings, by opposition to the simple summation function used in Section 2.1 for the phase identification with feeder per-phase current readings. The complex graph function that will be necessary to derive in order to identify the connection phases with voltage data, contrasts with the single-bus structure used before, and will be very useful to address the second problem.

The second problem will also require the use of a carefully designed function, $h(X)$ to capture the complex relationships between LV unbalanced loads and single-phase measurement quantities available, but will have to deal with a possibly more serious difficulty: the lack of data. LV data usually is available in very few locations and will not allow redundancy. The estimation problem will become ill-posed (the gain matrix G will be ill-conditioned) and, therefore, require regularization to be solved.

Until now, we have always dealt with overdetermined problems. The number of equations has always outnumbered the unknowns. When it did not, we have removed unknowns – such as in the case of SE in which we have estimated just part of the state vector – or inserted equations by adding pseudo-measurements, this way making the inversion problem well-posed[21]. This is a good enough approach when the number of necessary equations added to make the problem well-posed is not the majority of the equations needed. When the number of pseudo-measurements is large in comparison with the other measurements available, the approach itself needs to be questioned. And the question seems to rely upon the way pseudo-measurements should be used?

Let us illustrate the idea of ill-posed problems with the photo of Fig. 39 in which a column capital and its shade are represented. The problem of inferring about the column properties from the represented shade is an inverse problem and it is ill-posed because one is asked to use lower dimensional data (the 2D shade) to infer about the higher dimensional properties of the capital (a 3D object). Without additional knowledge on the column being Corinthian and other kind of information, one could hardly infer about the properties of the column from the data on its shade. No matter how much data one would get on that shade. One possible way to infer about the capital is to expect that it is a standard Corinthian capital and use the shade to change such expected standard as little as possible in order to match the data on the shade. This is an analogy to regularization – the second term used to classify the kind of problems addressed in this chapter.

Figure 39: Photograph from McGoodwin's book *Architectural Shades and Shadows* illustrating a Corinthian column capital and its shade.

Back to SE, to regularize the ill-conditioned estimation problem we will have to use pseudo-measurements of LV historical loads. The only reliable information on such loads is their expected ("standard") magnitude, as loads differ in magnitude quite a lot. So, it will be the magnitude that we will to use to regularize the ill-conditioned SE problems in order to solve them properly.

[21]Well-posed problems (in the Hadamard sense) are problems for which a solution exists, is unique, and changes continuously with the problem's initial conditions.

Properly here means without using expected magnitudes as ordinary measurements whose (lack of) precision one weights in WLS to estimate LV states.

In the sections that follow, we will look into possible solutions to LV grid problems with emphasis on solutions to the phase identification and the state estimation cases.

4.1 Phase identification with three-phase voltage metering data

As we have said earlier in Chapter 2, the topology model of LV grids usually lacks reliable phase connectivity information of single-phase connected customers. We have addressed the problem of estimating phase connectivity information based on smart metering data on customer's consumption with measurements on per-phase aggregate energy totals assuming these were available at substation sites with the same resolution. When such per-phase measurements are available, the problem of phase identification can be formulated as a multi-linear regression classification problem and be solved efficiently without domain specific prior knowledge on power system analysis. When not available, the going gets a bit tougher.

Unfortunately, per-phase aggregate energy measurements are usually unavailable at the feeder or transformer levels, this way making the massive estimation of phase connectivity impossible without information on grid specific properties. Instead, voltage metering capabilities (sometimes also μPMU[22]) are sometimes available from three-phase meters deployed at customer premises.

When voltage metering data are available, the phase-identification problem can be stated in a very similar way as before. The difference is that the per-phase readings are now voltage values. We recall the formulation stated before in the following:

Consider N customers whose connection phase assumes one out of three possible labels a, b, or c and estimate the correct customer-to-label assignment based on M readings whose per-phase values are a *function* of the corresponding phase-connected customer individual energy measurements.

Voltage is a variable that can be used to describe the system state and therefore be used to infer on loads connection phase. Note that the set of nodal voltage magnitudes and arguments fully describes the system state as all the currents can be computed with

[22]The PMU acronym refers to Phasor Measurement Unit. PMU's, also called synchrophasors, measure the angle between voltages at different physical locations on a grid. PMU's have been used in transmission grids. Distribution grids have much tinier angle differences – too small, and changing too rapidly, to resolve with traditional PMU's. Synchrophasor measurement technologies adapted to distribution grids are called micro-PMU or μPMU. This technology provides super-fast, accurate, network-wide synchronised measurements of phasors and amplitudes for voltage and current up to 512 samples per nominal cycle (corresponds to 25,600 samples/s @ 50Hz and 30,720 samples/s @ 60Hz).

the voltages for known two-port/multi-terminal impedance models of the lines. However, unlike feeder currents, the nodal voltage dependence on loads and their connection phases is a complex one. Dependence involves complex relationships between the phase currents and the neutral current (at the vertex/nodal level) and complex topology relationships between all such currents (at the edge/branch level).

With voltages data alone, the formulation of the phase identification problem now requires more than basic prior knowledge on power system analysis to be carried out. Before embracing into formulating the identification problem, we now need to understand how to model the four-wire three-phase LV circuits. Also, the solution now requires more than just data on loads and per-phase readings to be obtained. It requires information on the grid topology and on the two-port/multi-terminal impedances of its lines.

Let us see how we could use such information to express per-phase voltage y^p as a *function* of the N per-phase individual energy measurements x_i and corresponding binary phase assignments, β_i^p. Let us recall the function used before, in (3), in which the readings y^p were expressed simply as a summation of the energy measurements x_i:

$$y^p(k) \approx \sum_{i=1}^{N} \beta_i^p x_i(k), \quad p \in \{a, b, c\}$$

Could the three phase voltage readings in a given bus, y^p, be also expressed as a summation *function* of the N per-phase individual energy measurements x_i and corresponding binary phase assignments, β_i^p? Let us look at Fig. 40 and try to figure out the answer.

The voltage difference between buses, say i and j, can be expressed on the currents flowing into branch ij by making use of the Ohm's law and of the information on the branch impedance $z_{ij} = r_{ij} + ix_{ij}$, where the non-subscript "i" symbol refers to the imaginary number, $i = \sqrt{-1}$. The voltage difference could be expressed in matrix form as $dV_{ij}^{abcn} = Z_{ij}^{abcn} I_{ij}^{abcn}$, where

$$dV_{ij}^{abcn} = \begin{pmatrix} dv_{ij}^{ag} \\ dv_{ij}^{bg} \\ dv_{ij}^{cg} \\ dv_{ij}^{ng} \end{pmatrix}$$

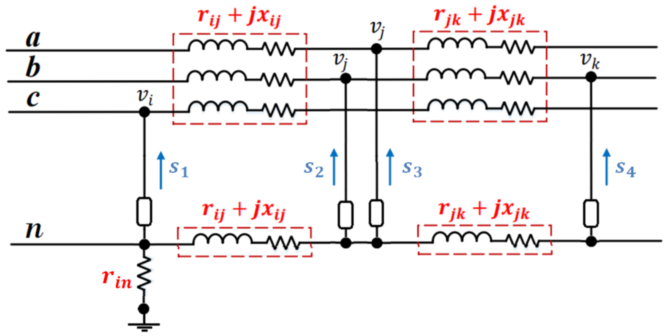

Figure 40: Four-wire (plus ground) three-phase representation of an LV grid section (two branches and three nodes, i, j, and k) explored under an unbalanced loading situation.

$$I_{ij}^{abcn} = \begin{pmatrix} i_{ij}^{a} \\ i_{ij}^{b} \\ i_{ij}^{c} \\ i_{ij}^{n} \end{pmatrix}$$

$$Z_{ij}^{abcn} = \begin{pmatrix} z_{ij} & 0 & 0 & 0 \\ 0 & z_{ij} & 0 & 0 \\ 0 & 0 & z_{ij} & 0 \\ 0 & 0 & 0 & z_{ij} \end{pmatrix}.$$

The branch impedance matrix Z_{ij}^{abcn} can be assumed to be approximately diagonal as mutual inductances are usually negligible in LV, i.e., currents in a given phase are considered not to cause significant voltage changes in the other phases or the neutral conductors. Also, its diagonal elements can be assumed to be identical as the neutral conductor usually is identical to the phase conductors in most LV networks[23].

Assuming that the voltage changes in the four-wire system of Fig. 40 can be expressed by the four dimensional Ohm law, i.e., $dV_{ij}^{abcn} = Z_{ij}^{abcn} I_{ij}^{abcn}$, then if there is no significant conductance from the neutral to the ground[24], i.e., $r_{in} \gg z_{ij}$, the neutral cur-

[23] In some urban networks of tetra-polar underground cables, smaller cross section areas are sometimes used for the neutral conductor, making the former assumption invalid.

[24] European kind of LV networks use TT or TN-C earthing systems. In TT, the neutral (N) conductor is grounded at the substation transformer only. In TNC and TNC-S, the neutral (N) is also grounded at the substation transformer, being used as the sole protective earth (PE) between the substation

rent i_{ij}^n can be expressed as the negated sum of the other three currents phasors, as follows:

$$i_{ij}^n = -i_{ij}^a - i_{ij}^b - i_{ij}^c. \tag{57}$$

The neutral current expression (57) can be generalized for any branch of a radial grid. The generalization is easy to accept by imagining a meta node that circumvents all the nodes downstream the receiving-end of a given branch, and noticing that the KCL imposes the equality of the sum of all currents flowing into the meta node. See Fig. 41 for an illustration of such imposition.

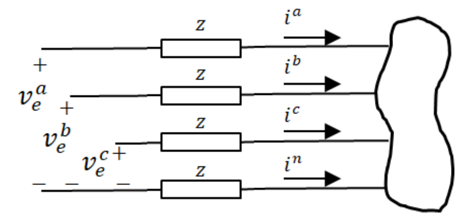

Figure 41: Illustration of equation (57) in which the KCL imposes the sum of all current phasors to be zero.

If (57) holds, then one may rewrite the phase-to-neutral voltage changes as a three dimensional Ohm law, as follows:

$$dV_{ij}^{abc} = Z_{ij}^{abc} I_{ij}^{abc} \tag{58}$$

Where,

$$dV_{ij}^{abc} = \begin{pmatrix} dv_{ij}^a \\ dv_{ij}^b \\ dv_{ij}^c \end{pmatrix}$$

$$I_{ij}^{abc} = \begin{pmatrix} i_{ij}^a \\ i_{ij}^b \\ i_{ij}^c \end{pmatrix}$$

$$Z_{ij}^{abc} = \begin{pmatrix} 2z_{ij} & z_{ij} & z_{ij} \\ z_{ij} & 2z_{ij} & z_{ij} \\ z_{ij} & z_{ij} & 2z_{ij} \end{pmatrix} = z_{ij} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Note that according to the four dimensional Ohm law, the phase-to-neutral voltages can be expressed in the wire-to-ground phasor variables as $z_{ij}i^a - z_{ij}i^n = z_{ij}(i^a - i^n)$, which using (57) results in $z_{ij}(2i^a + i^b + i^c)$. Consequently, the phase-to-neutral quantities $dv^a = dv^{ag} - dv^{ng}$ can be expressed as in (58). Fig. 42 represents the voltages change at the receiving-end of a branch, caused by a single-phase load current and corresponding neutral return, as computed with (58) for balanced symmetrical sending-end voltages.

and the entry point to the customer facilities – we call it a combined PEN conductor –, and separated into a dedicated neutral conductor (N) and a protective earth conductor (PE) conductor beyond the entry point, indoor customer facilities only. The PEN conductor might sometimes be earthed in a few poles of the overhead grid, but the earthing electrodes are usually bad enough for us to ignore earth as a current returning alternative path.

Assuming the distribution grid is radial, all the branch currents can be obtained by defining a meta node over the set of nodes downstream the receiving-end of each branch. Thus, per-phase branch currents can be estimated by summing the per-phase load currents of every node in the set of nodes downstream the receiving-end of each branch. Let us call N_{ij} the set of nodes downstream the receiving-end of branch ij. The per-phase p currents i_{ij}^{p} can be obtained by:

$$i_{ij}^{p} = -\sum_{n \in N_{ij}} (s_n^p / v_n^p)^{\star}, \quad p \in \{A, B, C\} \tag{59}$$

Where s_n^p refers to the complex power injection in phase p of bus n, v_n^p refers to the corresponding phase-to-neutral voltage, and the star ($\star$) symbol refers to the complex conjugate operator.

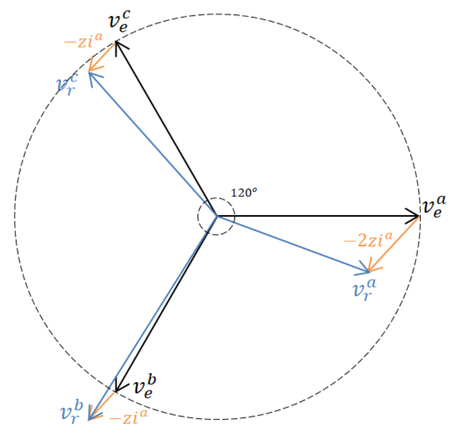

Figure 42: Phasor representation of the voltage changes cause by a single-phase load current and corresponding neutral return, as computed with (58) for balanced symmetrical sending-end voltages

If we aim to express (58) on the currents computed with (59), we could obtain the voltage approximated changes in a given branch ij with:

$$dV_{ij}^{abc} = Z_{ij}^{abc}[-\sum_{n \in N_{ij}} (s_n^{abc} / v_n^{abc})^{\star}] = -Z_{ij}^{abc}[\sum_{n \in N_{ij}} i_n^{abc}] \tag{60}$$

and sum branch voltage changes dV_{ij} along the graph path to the source bus to compute the voltage deviation dV_k in a given bus k as follows:

$$dV_k^{abc} = -\sum_{ij \in A_{ks}} [Z_{ij}^{abc} \sum_{n \in N_{ij}} i_n^{abc}] \tag{61}$$

where A_{ks} is the set of branches in the path from node k to the source node s.

The above double summation can now be rewritten to isolate nodal injected currents, as follows:

$$dV_k^{abc} = -\sum_{n \in N} [i_n^{abc} \sum_{ij \in A} Z_{ij}^{abc}] = -\sum_{n \in N} [\sum_{ij \in A} Z_{ij}^{abc}] i_n^{abc} \tag{62}$$

where A is the set of branches in the path from node n to the source node that contain the path from k to such source, i.e., $A = A_{ns} \cap A_{ks}$.

We may designate the sum of impedances $\sum_{ij \in A} Z_{ij}^{abc}$ by W_{kn} and express the voltage deviations as a linear combination of the complex injected currents, as follows:

$$dV_k^{abc} = -\sum_{n \in N} W_{kn} i_n^{abc} \tag{63}$$

which can be expressed in matrix form as follows:

$$dV_k^{abc} = \begin{pmatrix} W_{k1}i_1 & W_{k2}i_2 & \dots & W_{kn}i_n \end{pmatrix} \begin{pmatrix} \gamma_1^{abc} \\ \gamma_2^{abc} \\ \vdots \\ \gamma_n^{abc} \end{pmatrix} \tag{64}$$

where i_1 is the load current in node 1, i_2 is the load current in node 2, etc., whose connection phase we do not know. These currents constants (not vectors). These are complex constants whose values are symmetric to the injected currents in each node. They are complex numbers because we need to account for loads' power factor, when known, but are not shifted according to their source voltage $\{a, b, c\}$, because we ignore the phase to which they are connected (remember that all this effort is to estimate such phases).

The information on the necessary shifting of each load current i is provided by the load's phase complex assignment γ_i, where:

$$\gamma_i^{abc} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha^2 \end{pmatrix} \begin{pmatrix} \beta_i^a \\ \beta_i^b \\ \beta_i^c \end{pmatrix}, \quad i = 1, \dots, n$$

where, $\alpha = e^{-i120^{\mathrm{o}}}$

Equation (64) can be written on real/binary assignments by shifting the columns of the impedance sums instead, i.e., by multiplying each sum W_{ki} by a diagonal matrix of shifts, D_α, as follows:

$$dV_k^{abc} = \begin{pmatrix} D_\alpha \circ W_{k1}i_1 & D_\alpha \circ W_{k2}i_2 & \dots & D_\alpha \circ W_{kn}i_n \end{pmatrix} \begin{pmatrix} \beta_1^{abc} \\ \beta_2^{abc} \\ \vdots \\ \beta_n^{abc} \end{pmatrix} \tag{65}$$

where,

$$D_\alpha = \begin{pmatrix} 1 & \alpha & \alpha^2 \\ 1 & \alpha & \alpha^2 \\ 1 & \alpha & \alpha^2 \end{pmatrix}$$

and the symbol "$\circ$" refers to the Hadamard (or element-wise) matrix product.

With M readings on per-phase voltages in a given node k, V_k, and M readings on unknown-phase loads (or load currents), i_n, the complex expansion of (65) can be written as a regression equation:

$$y = X\beta + \epsilon \tag{66}$$

and be solved by $\beta = (X^* X)^{-1} X^* y$ where X^* refers to the Hermitian transpose of X and,

$$y = \begin{pmatrix} dV_k^a(m) \\ dV_k^b(m) \\ dV_k^c(m) \\ \ldots \end{pmatrix}, \; m = 1, \ldots M$$

$$X = \begin{pmatrix} D_\alpha \circ W_{k1} i_1(m) & D_\alpha \circ W_{k2} i_2(m) & \ldots & D_\alpha \circ W_{kn} i_n(m) \\ \vdots & \vdots & \ldots & \vdots \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_i^a \\ \beta_i^b \\ \beta_i^c \\ \vdots \end{pmatrix}, \; i = 1, \ldots, n$$

Note that, as before, the solution of the regression equation will be an array of real numbers; not an array of binary labels, as required for identification of the customer phase connections.

To label the customers one needs to map each group of three entries of β into a binary vector of a single 1's value. This can be achieved by identifying the component whose real value is closer to the unitary value, as follows:

$$\beta_i^p = \begin{cases} 1 & \Leftarrow arg \; min_{\{A,B,C\}} \; \|Re(\beta(3(i-1)+1, 3i)) - 1\| = p \\ 0 & \Leftarrow \text{otherwise} \end{cases}, \quad i = 1, \ldots, n. \tag{67}$$

As written, the regression equation requires voltages deviations, dV^{abc}, to be known in both their magnitude and argument. That requires two things. One is that the voltage be also measured at the source node (substation busbar). The other is that the voltage be measured at the busbar and the feeder node in both its magnitude and argument, which requires the installation of μPMUs at both grid locations. If only voltage magnitude meters are available, then the equation needs to be rewritten on its absolute terms. This can be done in different manners. In the following example we will illustrate how to do that without dramatically compromising labelling accuracy.

```
s = 0.0450    0.0150    0.0470    0.0330
    0.0250    0.0150    0.2480    0.0330
    0.0970    0.0250    0.3940    0.0330
    0.0700    0.0490    0.0200    0.4850
    0.1250    0.0460    0.0160    0.1430
    0.2900    0.0270    0.0160    0.0470
    0.2590    0.0150    0.0170    0.0200
    0.2590    0.0160    0.0280    0.0160
    0.4420    0.0160    0.0500    0.0170
    0.2010    0.0230    0.0460    0.0160
    0.2060    0.0490    0.0220    0.0240
    0.1300    0.0470    0.0160    0.0490
    0.0460    0.0260    0.0170    0.0480
```

```
v = 0.9987 - 0.0048i  -0.4983 - 0.8684i  -0.4939 + 0.8596i
    0.9729 - 0.0160i  -0.5114 - 0.8738i  -0.5074 + 0.8546i
    0.9561 - 0.0244i  -0.5189 - 0.8775i  -0.5133 + 0.8491i
    1.0477 - 0.0355i  -0.4509 - 0.8976i  -0.4010 + 0.7933i
    1.0158 - 0.0100i  -0.4830 - 0.8724i  -0.4657 + 0.8443i
    1.0097 - 0.0070i  -0.4892 - 0.8706i  -0.4777 + 0.8517i
    1.0060 - 0.0054i  -0.4928 - 0.8699i  -0.4848 + 0.8557i
    1.0042 - 0.0057i  -0.4939 - 0.8698i  -0.4863 + 0.8560i
    1.0058 - 0.0100i  -0.4910 - 0.8734i  -0.4791 + 0.8494i
    1.0007 - 0.0054i  -0.4963 - 0.8685i  -0.4901 + 0.8578i
    1.0048 - 0.0028i  -0.4936 - 0.8648i  -0.4866 + 0.8590i
    1.0063 - 0.0032i  -0.4924 - 0.8655i  -0.4846 + 0.8578i
```

Figure 43: Matrix of customer power readings s and the corresponding noiseless per-phase voltage measurements v in bus k, in each of the 12(+1) time periods.

Four-customers, 12-periods Phase-id (2nd) Example: Take the training set of data points depicted in Fig. 43 where s represents the apparent power readings and v the corresponding complex per-phase voltage measurements obtained during 12 periods of 15 minutes (plus the final reading) for four single-phase connected customers $\{1, 2, 3, 4\}$ whose connection phases are unknown. Assume that voltages are measured at the terminal feeder node k and use information on the feeder impedances to determine the customer connection phases. The topology that maps the customers into the feeder buses is depicted in Fig. 44.

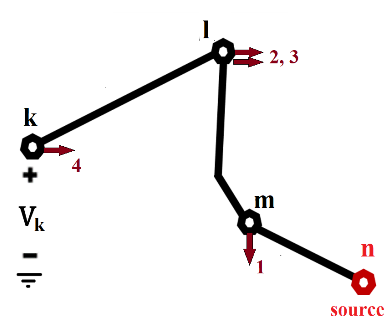

Figure 44: Topology and monitored variables of the LV feeder used in the phase id example.

Assuming that source voltages V_n^{abc} are known and that feeder impedances can be accurately determined, one may use (65) to express voltage deviations at the terminal node, V_k^{abc} as a function of the phase connection attributes of feeder customers, β_i^{abc}, as:

$$dV_k^{abc}(m) = \begin{pmatrix} D_\alpha \circ W_{k1} i_1(m) & D_\alpha \circ W_{k2} i_2(m) & D_\alpha \circ W_{k3} i_3(m) & D_\alpha \circ W_{k4} i_4(m) \end{pmatrix} \begin{pmatrix} \beta_1^{abc} \\ \beta_2^{abc} \\ \beta_3^{abc} \\ \beta_4^{abc} \end{pmatrix} \tag{68}$$

With,

$$W_{k1} = z_{nm} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$W_{k2} = W_{k3} = (z_{nm} + z_{ml}) \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$W_{k4} = (z_{nm} + z_{ml} + z_{lk}) \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

Assuming that source voltages are balanced, symmetrical, and unitary, i.e., $V_n^{abc} = \begin{pmatrix} 1+0i & -0.5-0.866i & -0.5+0.866i \end{pmatrix}$, and that branch impedances are,

$$\begin{aligned} z_{nm} &= 0.0250 + 0.0125i \\ z_{ml} &= 0.0375 + 0.0175i \\ z_{lk} &= 0.0500 + 0.0250i \end{aligned}$$

Then, results of the regression equation are:

$$\begin{pmatrix} \beta_1^{abc} \\ \beta_2^{abc} \\ \beta_3^{abc} \\ \beta_4^{abc} \end{pmatrix} = \begin{pmatrix} 0.0001 + 0.0027i \\ 0.0000 + 0.0009i \\ 0.9998 - 0.0132i \\ -- \\ 0.0001 + 0.0064i \\ 1.0001 + 0.0001i \\ 0.0032 + 0.1228i \\ -- \\ 0.9997 - 0.0222i \\ -0.0000 - 0.0011i \\ -0.0000 + 0.0078i \\ -- \\ -0.0000 - 0.0003i \\ -0.0000 + 0.0017i \\ 0.9983 - 0.0608i \end{pmatrix}$$

Whose real values can be used to easily assign the correct phase connection with (67), as follows:

$$Re(\beta_1^{abc}) = \begin{pmatrix} 0.0001 \\ 0.0000 \\ 0.9998 \end{pmatrix} \rightarrow p_1 := c$$

$$Re(\beta_2^{abc}) = \begin{pmatrix} 0.0001 \\ 1.0001 \\ 0.0032 \end{pmatrix} \rightarrow p_2 := b$$

$$Re(\beta_3^{abc}) = \begin{pmatrix} 0.9997 \\ 0.0000 \\ 0.0000 \end{pmatrix} \rightarrow p_3 := a$$

$$Re(\beta_4^{abc}) = \begin{pmatrix} 0.0000 \\ 0.0000 \\ 0.9983 \end{pmatrix} \rightarrow p_3 := c$$

If we add noise to the per-phase voltages, results may be harder to interpret. See the output obtained for noisy voltage readings obtained for $\epsilon \sim \mathcal{N}(0, \sigma^2)$, with $\sigma = 0.005pu$ – a large voltage error.

The correct phase connection can be assigned with (67) over the obtained results. Yet, some phase connections seem now to be harder to identify, e.g., node 2 connection to phase b in bus l is not easy to identify – the real part of the corresponding β coefficient is quite distant from the unit value.

$$\begin{pmatrix} \beta_1^{abc} \\ \beta_2^{abc} \\ \beta_3^{abc} \\ \beta_4^{abc} \end{pmatrix} = \begin{pmatrix} 0.5657 - 0.2793i \\ -0.0256 - 0.2120i \\ 1.1350 + 0.1335i \\ -- \\ -1.8501 + 1.3485i \\ 1.8207 + 0.2992i \\ -0.8537 - 0.5390i \\ -- \\ 0.9667 - 0.0361i \\ -0.1415 + 0.0488i \\ -0.0457 + 0.0763i \\ -- \\ 0.1623 - 0.1292i \\ -0.0309 - 0.0260i \\ 1.1172 - 0.0400i \end{pmatrix}$$

Errors in μPMU voltage measurements are usually small – much smaller than the error just introduced. Also, possible difficulties in grid modelling, namely in obtaining accurate enough feeder branches impedances values, do not usually cause labelling difficulties.

More serious difficulties may yet appear when μPMU are not available and regular voltage magnitude meters need to be used. The inaccuracies introduced by assuming known voltage arguments are sometimes significant. Recall Fig. 42 and note that voltage argument shifts caused by unbalanced load currents are very different from phase to phase. Not knowing the arguments of the voltage measurements, i.e., assuming these arguments are similar, introduces an error that may cause labelling inaccuracies. In the following, we show the result of taking y as the difference between the magnitudes of the source voltage and the measured voltage, i.e.:

$$dV_k^{abc} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha^2 \end{pmatrix} \left(\|V_n^{abc}\| - \|V_k^{abc}\|\right) \tag{69}$$

Results are presented numerically in the following; the data is shown graphically in Fig. 45.

$$\begin{pmatrix} \beta_1^{abc} \\ \beta_2^{abc} \\ \beta_3^{abc} \\ \beta_4^{abc} \end{pmatrix} = \begin{pmatrix} -0.5469 - 0.1819i \\ 0.3085 + 0.4824i \\ 1.0330 - 0.6978i \\ -- \\ 0.3103 + 0.5395i \\ 1.1624 - 0.7387i \\ -0.5803 - 0.2292i \\ -- \\ 1.0533 - 0.6726i \\ -0.5318 - 0.2047i \\ 0.2839 + 0.4907i \\ -- \\ -0.5432 - 0.2123i \\ 0.2552 + 0.5078i \\ 1.0581 - 0.6719i \end{pmatrix}$$

Again, the phase connectivity can be correctly assigned with (67) over the results obtained by regression with voltage magnitudes only.

Yet, if we would have added noise to the voltage magnitude measurements (as we did before), node 2 connection to phase b in bus l would again be hard to identify – the real part of the β

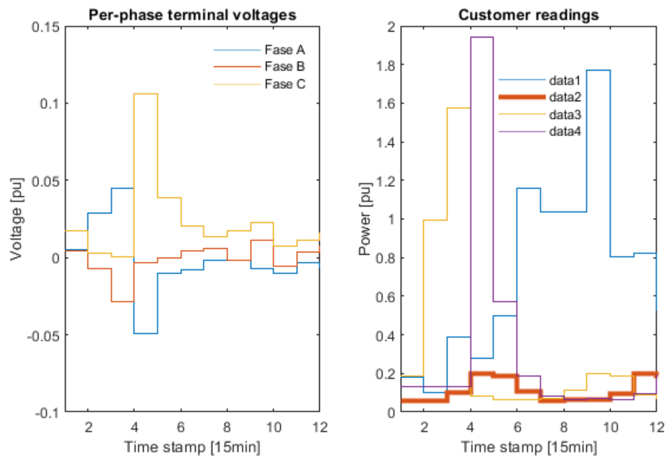

Figure 45: Chronological representation of per-phase voltage drops, dV_k^{abc} as computed with (69) (left-hand side) and the corresponding customer power output, $s_i, i = 1, \ldots, 4$ (right-hand side) for which phase labels (A, B, C) need to be assigned.

coefficient would again be distant from the unit value. That is somehow expected as the load in node 2 is much smaller that the other three loads (see the red plotted profile in Fig. 45 and compare it with the other three profiles). Being always much smaller, the phase connection impacts of such load will be much harder to correlate with the voltage changes measured in the terminal node. Remember that covariances are given by $(X^T X)^{-1}\sigma^2$ and that X is proportional to the loads.

In the figure, when load peaks, it seems easy to correlate voltage changes with such peaks. For instance, the high voltage drop in period 4-5 (the rise shown in Phase C profile change) makes us suspect from the purple load being connected to phase C. The purple load is load 4. The peaking of the yellow load profile in the periods just behind and the rise in Phase A profile change also points out to load 3 being probably connected to phase A. As for the other loads, labelling seems hard to make by visual inspection. Reasons might be that load 1 is electrically too close to the source, and that load 2 is always too small, to impact significantly on the feeder terminal voltage.

The difficulty in labelling the connection phase of small loads is however more formal than practical. If the load is small, the errors caused by erroneous labelling of such load are negligible, making the correct labelling unnecessary in practice.