




Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

Signal and Information  
Processing Laboratory  
Institut für Signal- und  
Informationsverarbeitung 

---

Winter Semester 2001/2002

Prof. Dr. H.-A. Loeliger

Post-Diploma Project

# Kalman Filters, Factor Graphs, and Electrical Networks

Pascal O. Vontobel

---

Advisor: Prof. Dr. Hans-Andrea Loeliger

ISI-Number: INT/200202



# Abstract

This post-diploma project highlights connections between Kalman filters, factor graphs, and electrical networks.

Generally speaking, we are interested in the blockwise maximum a-posteriori estimation of random variables based on measurements of some related random variables. Such problems can very conveniently be stated using factor graphs that represent the joint probability density of all occurring random variables. Having found a factor graph, one usually uses the max-product algorithm to get estimates.

When one devises an electrical network that solves the same estimation problem, then one realizes that there is a topological one-to-one relationship between the factor graph and the electrical network. As a key example we consider the factor graph of the Kalman filter and the electrical network which solves the Kalman filter problem. It must be emphasized that the electrical networks always find the correct solution (in the context of convex problems as studied in this report) whereas the max-product algorithm guarantees to give the correct results only when the factor graph is a tree.

It turns out, that results from electrical network theory like Tellegen's theorem, Green's reciprocity theorem, and dualization have corresponding results in estimation and optimization theory. Within our framework it is possible to switch back and forth between factor graphs and electrical networks. This approach is much more related to the results in the book by Dennis [1] and the diode decoder by Davis and Loeliger [2] than to the analog decoder by Loeliger *et al.* [3]. Whereas the latter is an implementation of an algorithm (the sum-product algorithm) using analog electrical circuits, the former two implement the problem itself, "leaving it to nature" to solve the problem.<sup>1</sup>

Other topics included in this report range from the fact that simplifying an electrical network corresponding to a factor tree is the max-product algorithm, a matrix-based approach for updating the messages in a factor graph of jointly Gaussian random variables, a probabilistic interpretation of electrical networks given in the book by Mead [4], and several duality results.

---

<sup>1</sup>Therefore also the title of [2]: "A *Nonalgebraic* Maximum Likelihood Decoder for Trellis Codes" (emphasis added).



# Acknowledgments

First of all, I would like to thank my advisor Andi Loeliger for his constant support of my ideas and for giving me good advice during the whole project. The joint discussions were always a big pleasure.

A warm thanks also goes to Sanjoy Mitter from MIT who, while visiting our lab in the summer of 2001, mentioned the book by Dennis [1]. This book motivated and inspired this project in a very fundamental way.

I would also like to thank Dani Lippuner from Siemens AG (formerly ISI, ETHZ) for discussions about the subject and for pointing out to me the paper by Carter [5].

This work was supported by grant TH-16./99-3.

Zurich, April 10, 2002

Pascal O. Vontobel

In Sec. 1.6 we corrected two errors in the formulas. (July 10, 2002)

Eq. (C.10) has been corrected. (July 17, 2002)

Thanks to a remark by Patrick Merkli from ETH Zurich, Eqs. (4.28) to (4.33) have been corrected. (February 27, 2003)

The resistances in Fig. 2.12 have been changed from  $1/\sigma_{U[k]}^2$  and  $1/\sigma_{W[k]}^2$  to  $\sigma_{U[k]}^2$  and  $\sigma_{W[k]}^2$ , respectively. The wiring has also been changed slightly in that figure. (March 22, 2007)



# Contents

<b>1</b>	<b>Introduction and Basics</b>	<b>1</b>
1.1	INTRODUCTION . . . . .	1
1.2	FORNEY-STYLE FACTOR GRAPHS (FFGs) . . . . .	2
1.3	THE SUM-PRODUCT ALGORITHM (SPA) . . . . .	4
1.4	THE MAX-PRODUCT ALGORITHM (MPA) . . . . .	5
1.5	OTHER ALPHABETS . . . . .	5
1.6	DIRAC-DELTA DISTRIBUTIONS AS LIMITING DISTRIBUTIONS . . . . .	6
1.7	THE CHARACTERISTICS OF THE ELECTRICAL NETWORK ELEMENTS THAT ARE USED . . . . .	6
<b>2</b>	<b>Simple Motivating Examples</b>	<b>11</b>
2.1	EXAMPLE 1 (SUMMATION FUNCTION NODE) . . . . .	11
2.2	EXAMPLE 2 (SUMMATION-FUNCTION AND EQUALITY-FUNCTION NODES) . . . . .	13
2.3	EXAMPLE 3 (NON-GAUSSIAN RANDOM VARIABLES) . . . . .	14
2.4	EXAMPLE 4 (A SIMPLE KALMAN FILTER) . . . . .	16
<b>3</b>	<b>From a Forney-style Factor Graph to an Electrical Network</b>	<b>21</b>
3.1	THE OVERALL LAGRANGIAN . . . . .	21
3.2	EDGES . . . . .	22
3.2.1	EDGES BETWEEN TWO CONSTRAINT FUNCTION NODES . . . . .	22
3.2.2	EDGES BETWEEN A LEAF-FUNCTION NODE AND A CONSTRAINT-FUNCTION NODE . . . . .	23
3.2.3	EDGES BETWEEN TWO ONE-DEGREE FUNCTION NODES . . . . .	23
3.3	HALF-EDGES . . . . .	23
3.4	SUMMATION-FUNCTION NODE . . . . .	24
3.5	EQUALITY-FUNCTION NODE . . . . .	25
3.6	SCALING-FUNCTION NODE . . . . .	25
3.7	PUTTING IT ALL TOGETHER . . . . .	26
<b>4</b>	<b>The Max-Product Algorithm and the Sum-Product Algorithm</b>	<b>29</b>
4.1	EXAMPLES . . . . .	30
4.1.1	EQUALITY-FUNCTION NODE (EXAMPLE) . . . . .	30
4.1.2	SUMMATION-FUNCTION NODE (EXAMPLE) . . . . .	30
4.2	THE GENERAL CASE . . . . .	30
4.2.1	LEAF FUNCTION NODE (THE GENERAL CASE) . . . . .	31
4.2.2	EQUALITY-FUNCTION NODE (THE GENERAL CASE) . . . . .	32

4.2.3	SUMMATION-FUNCTION NODE (THE GENERAL CASE) . . . . .	33
4.2.4	SCALING-FUNCTION NODE (THE GENERAL CASE) . . . . .	35
4.2.5	CONCLUDING REMARKS (THE GENERAL CASE) . . . . .	36
4.3	MATRIX-METHOD FOR UPDATING MESSAGES . . . . .	37
4.3.1	EXAMPLE (EQUALITY-FUNCTION NODE) . . . . .	37
4.3.2	THE GENERAL CASE . . . . .	38
4.3.3	THE MATRIX-METHOD FOR UPDATING MESSAGES OF THE SCALAR KALMAN FILTER . . . . .	39
4.3.4	THE HIGHER-DIMENSIONAL CASE . . . . .	39
4.3.5	THE MATRIX-METHOD FOR UPDATING MESSAGES OF THE VECTOR KALMAN FILTER . . . . .	40
4.3.6	RENORMALIZATION . . . . .	41
4.4	SOME CONCLUSIONS ABOUT THE SCORE FUNCTION . . . . .	43
<b>5</b>	<b>Miscellaneous Topics</b> . . . . .	<b>45</b>
5.1	NONLINEAR RESISTORS AND A DICTIONARY BETWEEN PROBABILITY DEN- SITY FUNCTIONS AND $I(U)$ -CHARACTERISTICS . . . . .	45
5.2	THE VECTOR CASE . . . . .	45
5.3	THE ERROR COVARIANCE MATRIX IN THE CASE OF JOINTLY GAUSSIAN RANDOM VARIABLES . . . . .	46
5.3.1	INTRODUCTORY EXAMPLE . . . . .	46
5.3.2	THE GENERAL CASE . . . . .	47
5.4	EXAMPLES FROM THE BOOK BY MEAD . . . . .	52
5.5	CAPACITORS AND INDUCTORS . . . . .	53
5.5.1	COMPLEX RANDOM VARIABLES . . . . .	53
5.6	GRADIENT SEARCH . . . . .	54
5.7	INTERPRETATION OF THE DIODE-DECODER . . . . .	56
5.7.1	A SIMPLE EXAMPLE . . . . .	56
5.7.2	CANONICAL TRELLIS: THE GENERAL CASE . . . . .	57
5.7.3	GENERAL TRELLIS: THE GENERAL CASE . . . . .	58
<b>6</b>	<b>Primal-Dual FFGs and Multiports</b> . . . . .	<b>61</b>
6.1	THE PRIMAL AND THE DUAL PROBLEM . . . . .	61
6.1.1	THE PRIMAL PROBLEM . . . . .	61
6.1.2	THE DUAL PROBLEM . . . . .	62
6.1.3	PRIMAL FFG, DUAL FFG, AND PRIMAL-DUAL FFG . . . . .	63
6.2	TELLEGEN'S THEOREM AND MULTIPORTS . . . . .	65
6.2.1	TELLEGEN'S THEOREM . . . . .	65
6.2.2	MULTIPORTS . . . . .	66
6.2.3	ACTIVE/PASSIVE PART VERSUS NEUTRAL PART . . . . .	69
6.2.4	DUALIZING PARTS OF AN ELECTRICAL NETWORK . . . . .	70
6.3	OTHER TYPES OF DUALITIES . . . . .	72
6.3.1	CURRENT-VOLTAGE DUALITY . . . . .	72
6.3.2	FOURIER DUALITY OF FFGS . . . . .	72
6.3.3	PLANAR GRAPH DUALITY, PLANAR ELECTRICAL NETWORK DU- ALITY, AND MATROID DUALITY . . . . .	72



---

<b>7</b>	<b>Conclusions and Outlook</b>	<b>73</b>
<b>A</b>	<b>Some Facts from Estimation Theory</b>	<b>75</b>
A.1	A FIRST FAMILY OF ESTIMATORS . . . . .	75
A.1.1	CONDITIONAL EXPECTATION ESTIMATOR . . . . .	75
A.1.2	QUADRATIC COST FUNCTION . . . . .	75
A.1.3	ESTIMATOR WITH SMALLEST EXPECTED SQUARE ERROR . . . . .	75
A.1.4	BIAS-FREE MINIMUM-VARIANCE ESTIMATOR . . . . .	76
A.1.5	PROPERTIES OF THE FIRST FAMILY OF ESTIMATORS . . . . .	76
A.2	A SECOND FAMILY OF ESTIMATORS . . . . .	76
A.2.1	BLOCKWISE MAP ESTIMATOR . . . . .	76
A.2.2	SYMBOLWISE MAP ESTIMATOR . . . . .	76
A.2.3	UNIFORM COST ESTIMATOR . . . . .	77
A.2.4	UNIFORM COST FUNCTION (IN THE LIMIT) . . . . .	77
A.2.5	PROPERTIES OF THE SECOND FAMILY OF ESTIMATORS . . . . .	77
A.3	CRAMÉR-RAO BOUND . . . . .	77
A.4	JOINTLY GAUSSIAN RANDOM VARIABLES . . . . .	78
<b>B</b>	<b>The Node-Potentials Method</b>	<b>79</b>
B.1	FROM A LINEAR ELECTRICAL NETWORK TO A LINEAR SYSTEM OF EQUATIONS . . . . .	79
B.2	FROM A LINEAR SYSTEM OF EQUATIONS TO A LINEAR ELECTRICAL NETWORK . . . . .	80
<b>C</b>	<b>Lagrange Duality</b>	<b>81</b>
C.1	CONVEX FUNCTIONS AND THEIR CONJUGATE FUNCTION . . . . .	82
C.1.1	AN EXAMPLE . . . . .	83
C.2	DUALITY IN EQUALITY CONSTRAINED MINIMIZATION . . . . .	84
C.2.1	THE GENERAL CONVEX MINIMIZATION PROBLEM . . . . .	84
C.2.2	THE QUADRATIC MINIMIZATION PROBLEM . . . . .	85
C.3	ELECTRICAL NETWORKS WITH VOLTAGE SOURCES, CURRENT SOURCES AND RESISTORS . . . . .	86
C.4	MAX-PRODUCT ALGORITHM AND LAGRANGE DUALITY . . . . .	87
C.4.1	THE SUMMATION-FUNCTION NODE . . . . .	87
C.4.2	THE EQUALITY-FUNCTION NODE AND THE SCALING-FUNCTION NODE . . . . .	88
C.5	LAGRANGE DUAL OF SUBSPACE CONSTRAINTS . . . . .	89
C.6	FENCHEL'S INEQUALITY . . . . .	90
<b>D</b>	<b>Fourier Duality</b>	<b>93</b>
D.1	DUALIZATION OF THE SUMMATION-FUNCTION NODE . . . . .	93
D.2	DUALIZATION OF THE EQUALITY-FUNCTION NODE . . . . .	94
D.3	DUALIZATION OF SEVERAL FUNCTION NODES . . . . .	95
<b>E</b>	<b>Planar Graph and Electrical Network Duality</b>	<b>97</b>
E.1	PLANAR GRAPHS . . . . .	97
E.2	PLANAR ELECTRICAL NETWORKS . . . . .	98

---

<b>F</b>	<b>Singular Value Decomposition</b>	<b>101</b>
F.1	THE SINGULAR VALUE DECOMPOSITION . . . . .	101
F.2	THE MOORE-PENROSE GENERALIZED INVERSE . . . . .	101
<b>G</b>	<b>Tables</b>	<b>103</b>

# Chapter 1

## Introduction and Basics

### 1.1 Introduction

Our research started with the question: “How does an electrical network look like which solves the Kalman filter problem?” One motivation for this question was the book by Dennis [1], which gives the general solution of how to obtain an electrical network which solves a quadratic minimization problem using voltage and current sources, resistors, and DC-transformers. Dennis also considered problems with inequality constraints which in the resulting electrical networks turned out to be ideal diodes. But the main point in [1] was not to get the circuits themselves, but, based on the intuition of finding the voltages and currents of electrical networks, to come up with algorithms for solving the underlying abstract problem.

In this work one main aspect is the topology of the electrical networks that solve certain convex minimization problems. It turns out that not only for jointly Gaussian random variables (which corresponds to the quadratic minimization problem), there is a one-to-one correspondence between a Forney-style factor graph (FFG) and an electrical network.

Two other motivating subjects were on the one hand the electrical circuits that one can find in the book by Mead [4] and on the other hand the diode decoder as in the paper by Davis and Loeliger [2]. The concepts studied in this report can be applied to some of the circuits in [4] in order to give them a meaning in a probabilistic sense. We also try to reformulate the diode decoder in our present framework.

After having found the main results of this project, we saw that related work had been done by Carter [5] and Lippuner [6]. Carter and Lippuner realized that the calculation of the total resistance of serially and parallelly connected resistors are akin to the update rules for scalar Kalman filters. Lippuner also saw that the DC-transformer corresponds to the update rules for the estimate and variance in the case of a scaling-function node.

We finish this introduction by giving a brief overview over the report. The remaining sections of this chapter contain a short introduction into Forney-style Factor graphs (FFGs), the sum-product algorithm and the max-product algorithm, and discusses the electrical elements that will be used afterwards. Ch. 2 gives several examples how an electrical network can be devised that solves a blockwise maximum a-posteriori (MAP) problem that was stated using an FFG; from there it is not far anymore to the general rules as stated in Ch. 3. Then, Ch. 4 highlights the relation between the max-product/sum-product algorithms and network simplifications, whereas Ch. 5 discusses various topics in the realm of this project. In Ch. 6 we introduce a primal-dual FFG, which tries to unify a primal and its dual optimization problem, and relate

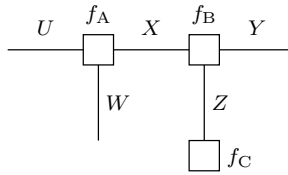


Figure 1.1: A Forney-style factor graph (FFG).

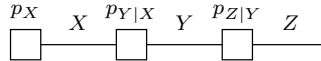


Figure 1.2: FFG of a Markov chain.

it to a multiport representation of the electrical networks obtained in Ch. 3. The main part closes with some conclusions and an outlook in Ch. 7.

Appendices A and B review the most important facts about estimation theory and about the node-potentials method, respectively. Appendices C, D, and E collect results from different types of duality: Lagrange duality, Fourier duality, planar graph duality, and planar electrical network duality. Finally, App. F deals with the singular value decomposition and the Moore-Penrose generalized inverse of a matrix, and App. G gives some tables summarizing some of the results of this report.

## 1.2 Forney-style Factor Graphs (FFGs)

A Forney-style factor graph (FFG) or “normal graph” [7] represents a factorization of a function of several variables<sup>1</sup>. For example, assume that some function  $f(u, w, x, y, z)$  can be factored as

$$f(u, w, x, y, z) = f_A(u, w, x) \cdot f_B(x, y, z) \cdot f_C(z). \quad (1.1)$$

This factorization is expressed by the graph of Fig. 1.1. In general, an FFG consists of nodes, edges, and “half edges”, where “half edges” are connected to only one node. The rules are as follows:

- There is a node for every factor.
- There is an edge (or half edge) for every variable.<sup>2</sup>
- The node representing some factor  $g$  is connected with the edge (or half edge) representing some variable  $X$  if and only if  $X$  is an argument of  $g$ .

Implicit in these rules is the assumption that no variable appears in more than two factors. We will see below that this condition is far less restrictive than might appear at first sight.

<sup>1</sup>This exposition follows closely [8].

<sup>2</sup>Note that upper case letters are used for edge labels, whereas lower case letters are used for configurations. This notation is inspired by the notation used in probability theory where upper case letters denote random variables and lower case letters denote realizations.

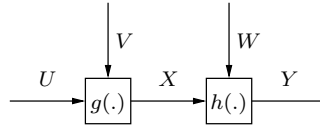


Figure 1.3: A block diagram.

The factors of the factorization expressed by the FFG are also called *local functions*; the overall function (i.e., the product of all local functions) is called the *global function*.

We will now rephrase some basic facts about factor graphs for FFGs; for more details, see [9] and [7]. We will first assume that all variables take values in finite sets; the modifications for continuous variables are given in Sec. 1.5.

In probability theory, factorizations of joint probability measures are expressions of independence. E.g., let  $X$ ,  $Y$ , and  $Z$  be discrete random variables with joint probability mass function  $p(x, y, z)$ . Then  $X, Y, Z$  form a Markov chain if and only  $p(x, y, z)$  can be factored as

$$p(x, y, z) = p(x) \cdot p(y|x) \cdot p(z|y). \quad (1.2)$$

This factorization is shown in Fig. 1.2. Upon removal of the edge  $Y$ , the graph falls into two disconnected components, with  $X$  and  $Z$  in different components, which expresses the conditional independence of  $X$  and  $Z$  given  $Y$ . It is easy to see that this generalizes to any FFG of a joint probability mass function: conditioned on the variables in any cut set of the graph, the variables in the two resulting components are independent.

A block diagram as in Fig. 1.3 may also be viewed as an FFG. A function block  $x = g(u, v)$  in the block diagram is then interpreted as representing the factor  $\delta(x - g(u, v))$ , where  $\delta$  is the Kronecker-delta function. When viewed as an FFG, the block diagram of Fig. 1.3 thus represents the function

$$\delta(x - g(u, v)) \cdot \delta(y - h(x, w)) = \begin{cases} 1, & \text{if } x = g(u, v) \text{ AND } y = h(x, w), \\ 0, & \text{else.} \end{cases} \quad (1.3)$$

In other words, the global function evaluates to 1 if and only if the values of all variables are consistent with the equations of the block diagram. Note that the arrows in the block diagram have no influence on its interpretation as an FFG.

As illustrated by these examples, an FFG can be used to express the structure of a “system” or “model”. In this context, the domain of the global function  $f$  is called the *configuration space*. A *configuration* is an element of the configuration space, i.e., a particular assignment of values to all variables. A configuration  $\omega$  is *valid* if  $f(\omega) \neq 0$ .

In a block diagram, we usually find also branching points as in Fig. 1.4(a). In an FFG, such branching points must be treated as factor nodes on their own, as is illustrated in Fig. 1.4(b). In doing so, there arise new variables ( $x_1$  and  $x_2$  in Fig. 1.4(b)) and a new factor

$$f_{=} (x, x_1, x_2) := \delta(x - x_1) \cdot \delta(x - x_2). \quad (1.4)$$

Note that, in every valid configuration, the new auxiliary variables have the same value as the original variable. By this device of variable “cloning”, it is always possible to enforce the condition that a variable may appear in at most two factors (local function).



Figure 1.4: A branching node in (a) becomes a replication node in (b).

One of the advantages of FFGs over factor graphs as defined in [9] is the clear distinction between external (“visible”) variables and internal (“latent” or “state”) variables: the former are represented by half edges, the latter by normal edges. Moreover, the operations of dividing a system into subsystems (“tearing”) and of “zooming” into the interior of some subsystem – both central to Willem’s system theory [10] – are naturally expressed in an FFG (cf. Figs. 2.9 and 2.10).

### 1.3 The Sum-Product Algorithm (SPA)

Another attraction of FFGs is that the sum-product algorithm (SPA) [9] [11] takes on a particularly simple form. Two messages are transmitted along each edge, one in each direction. (In practice, it often happens that only a subset of all these messages is actually needed.) Each message is, or represents, a *function* of the variable associated with that edge. Consider a node that represents some factor  $f(x_1, \dots, x_n)$ . The message  $\mu_{f \rightarrow X_k}$  out of this node along the edge  $X_k$  is the function

$$\mu_{f \rightarrow X_k}(x_k) = \sum_{x_1} \cdots \sum_{x_{k-1}} \sum_{x_{k+1}} \cdots \sum_{x_n} f(x_1, \dots, x_n) \cdot \mu_{X_1 \rightarrow f}(x_1) \cdots \mu_{X_{k-1} \rightarrow f}(x_{k-1}) \mu_{X_{k+1} \rightarrow f}(x_{k+1}) \cdots \mu_{X_n \rightarrow f}(x_n), \quad (1.5)$$

where  $\mu_{X_j \rightarrow f}$  is the message incoming on edge  $X_j$ . In words, the message  $\mu_{f \rightarrow X_k}$  is the product of the local function and all messages towards  $f$  along all edges except  $x_k$ , summed over all variables except  $X_k$ . In practice, the update rule (1.5) is often modified to include a scale factor.

The messages in the graph are computed, or iteratively recomputed, according to some schedule. As is well known, if the graph is cycle free, the (final) message out of some half edge representing some variable  $X$  is the marginal function

$$\mu(x) := \sum_{\omega: x \text{ fixed}} f(\omega), \quad (1.6)$$

where the sum goes over all configurations  $\omega$  with fixed  $x$ ; for details see [9]. Adopting the language from coding theory, we will call a decision (in the case of a discrete alphabet) or estimation (in the case of a continuous alphabet)

$$\hat{x} = \arg \max_x \mu(x) \quad (1.7)$$

a *symbolwise* maximum a-posteriori (symbolwise MAP) decision/estimation. If the graph is not a tree then the decision/estimation based on  $\mu(x)$  is not guaranteed to be the MAP symbolwise decision/estimation.

## 1.4 The Max-Product Algorithm (MPA)

Closely related to the sum-product algorithm is the max-product algorithm (MPA) [9] [11]. As for the SPA, two messages are transmitted along each edge, one in each direction, and each message is, or represents, a *function* of the variable associated with that edge. Consider a node that represents some factor  $f(x_1, \dots, x_n)$ . The message  $\mu_{f \rightarrow X_k}$  out of this node along the edge  $X_k$  is the function

$$\mu_{f \rightarrow X_k}(x_k) = \max_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n} f(x_1, \dots, x_n) \cdot \mu_{X_1 \rightarrow f}(x_1) \cdots \mu_{X_{k-1} \rightarrow f}(x_{k-1}) \mu_{X_{k+1} \rightarrow f}(x_{k+1}) \cdots \mu_{X_n \rightarrow f}(x_n), \quad (1.8)$$

where  $\mu_{X_j \rightarrow f}$  is the message incoming on edge  $X_j$ . In words, the message  $\mu_{f \rightarrow X_k}$  is the maximized product of the local function and all messages towards  $f$  along all edges except  $X_k$ , where the maximization is over all variables except  $X_k$ . In practice, the update rule (1.8) is often modified to include a scale factor.

The messages in the graph are computed, or iteratively recomputed, according to some schedule. As is well known, if the graph is cycle free, the (final) message out of some half edge representing some variable  $X$  is the summary function<sup>3</sup>

$$\mu(x) := \max_{\omega: x \text{ fixed}} f(\omega), \quad (1.9)$$

where the maximization goes over all configurations  $\omega$  with fixed  $x$ ; for details see [9]. Adopting the language from coding theory, we will call a decision (in the case of a discrete alphabet) or estimation (in the case of a continuous alphabet)

$$\hat{x} = \arg \max_x \mu(x) \quad (1.10)$$

a *blockwise* maximum a-posteriori (blockwise MAP) decision/estimation. This terminology stems from the fact that the  $x$ -component of  $\hat{\omega}$ , where

$$\hat{\omega} = \max_{\omega} f(\omega), \quad (1.11)$$

equals  $\hat{x}$  in Eq. (1.10) If the graph is not a tree then the decision/estimation based on  $\mu(x)$  is usually not the a MAP blockwise decision/estimation. Appendix A lists some well-known facts from estimation theory and gives conditions when the symbolwise (see Sec. 1.3) and the blockwise estimates are equal. Note that for jointly Gaussian random variables, the SPA and MPA are equivalent (for more details, see e.g. [8]).

## 1.5 Other Alphabets

In this report, we will be primarily interested in real or complex variables, or variables that are real or complex vectors. In this case, the Kronecker-delta in equations such as (1.3) and (1.4) should be replaced by the Dirac-delta distribution and the summations in (1.5) and (1.6) should be replaced by integrations.

---

<sup>3</sup>The term summary function is used in general for any message-passing algorithm, which all in a certain sense summarize some information contained in the graph.

## 1.6 Dirac-Delta Distributions as Limiting Distributions

It is often convenient to view Dirac-delta distribution as limits of Gaussian distributions.<sup>4</sup>

- Real case:

$$\delta(\mathbf{x}) = \lim_{\beta \rightarrow 0} \frac{1}{(2\pi\beta)^{n/2}} \exp\left(-\frac{1}{2\beta}\|\mathbf{x}\|_2^2\right), \quad (1.12)$$

- Complex case:

$$\delta(\mathbf{x}) = \lim_{\beta \rightarrow 0} \frac{1}{(\pi\beta)^n} \exp\left(-\frac{1}{\beta}\|\mathbf{x}\|_2^2\right). \quad (1.13)$$

The advantage of this representation is that it allows to rewrite densities that are products of a jointly Gaussian part and of Dirac-delta distribution terms as a single jointly Gaussian distribution. This simplifies sometimes the analysis (see e.g. Sec. 5.3). More detailed comments about this “unifying” approach can be found e.g. in [8].

## 1.7 The Characteristics of the Electrical Network Elements That Are Used

We now give a list of the electrical network elements that we will use. Such elements have a current  $I$  through them and a voltage  $U$  across them where the current arrow and the voltage arrow have the same direction. Only the last two elements are time-*dependent*, the others are all time-*independent*.

- **Ideal voltage source:** An ideal voltage source with parameter  $U_q$  has voltage-current characteristic  $U(I) = U_q$  independent of the current flowing through it. Fig. 1.5(left) shows the element and Fig. 1.5(right) the current-voltage characteristic.



Figure 1.5: Left: ideal voltage source with parameter  $U_q$ . Right: its current-voltage characteristic.

- **Real voltage source:** A real voltage source with parameters  $(U_q, R_q)$  has voltage-current characteristic  $U(I) = U_q + R_q I$ , or equivalently, the current-voltage characteristic  $I(U) = -U_q/R_q + U/R_q$ . Fig. 1.6(left) shows the element and Fig. 1.6(right) the current-voltage characteristic.
- **Ideal current source:** An ideal current source with parameter  $I_q$  has current-voltage characteristic  $I(U) = I_q$  independent of the voltage across it. Fig. 1.7(left) shows the element and Fig. 1.7(right) the current-voltage characteristic.

<sup>4</sup>Compared with [8] we have changed  $\beta$  to  $1/\beta$ .



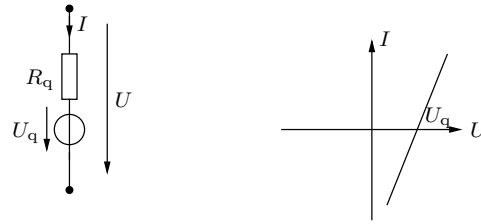


Figure 1.6: Left: real voltage source with parameters  $(U_q, R_q)$ . Right: its current-voltage characteristic (the slope is  $1/R_q$ ).

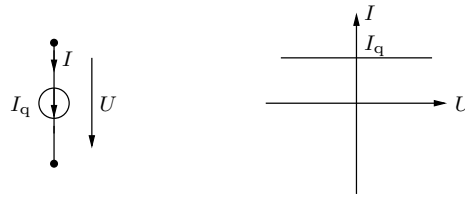


Figure 1.7: Left: ideal current source with parameter  $I_q$ . Right: its current-voltage characteristic.

- **Real current source:** A real current source with parameters  $(I_q, R_q)$  has current-voltage characteristic  $I(U) = I_q + U/R_q$ , or equivalently, the voltage-current characteristic  $U(I) = -R_q I_q + R_q I$ . Fig. 1.8(left) shows the element and Fig. 1.8(right) the current-voltage characteristic.

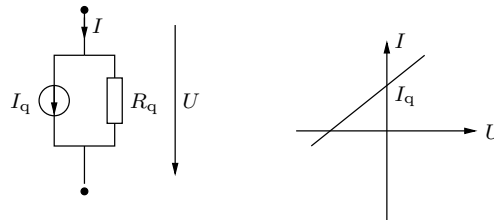


Figure 1.8: Left: real current source with parameters  $(I_q, R_q)$ . Right: its current-voltage characteristic (the slope is  $1/R_q$ ).

- **Linear resistor:** A linear resistor with parameter  $R$  has current-voltage characteristic  $I(U) = U/R$ , or equivalently, the voltage-current characteristic  $U(I) = RI$  (we allow the resistances to be also negative). Fig. 1.9(left) shows the element and Fig. 1.9(right) the current-voltage characteristic.
- **Nonlinear resistor:** A nonlinear resistor has current-voltage characteristic  $I(U)$ , or equivalently, the voltage-current characteristic  $U(I)$ . Fig. 1.10(left) shows the element and Fig. 1.10(right) an exemplary current-voltage characteristic. We assume that the product of current and voltage is at each operation point non-negative (meaning that this element is passive, i.e. dissipates energy), or zero (meaning that the element is

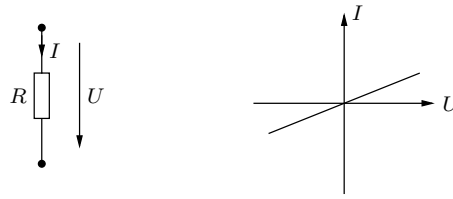


Figure 1.9: Left: linear resistor with parameter  $R$ . Right: its current-voltage characteristic (the slope is  $1/R$ ).

neutral).

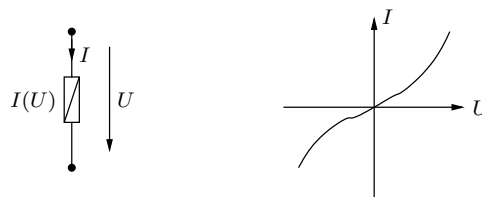


Figure 1.10: Left: nonlinear resistor. Right: exemplary current-voltage characteristic.

- **Nonlinear element:** A nonlinear element is described in the same way as a nonlinear resistor, but we allow the product of current and voltage to be positive, zero, or negative at different operation points meaning that the element can be active, neutral, and passive. Fig. 1.11(left and middle) show the element and Fig. 1.11(right) an exemplary current-voltage characteristic. Note that a nonlinear element whose current-voltage characteristic is monotonically increasing can be replaced by an ideal voltage source and a nonlinear resistor in series.

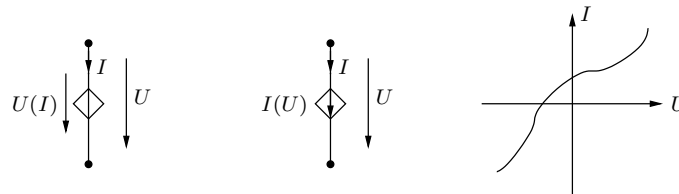


Figure 1.11: Left: non-linear element with voltage-current characteristic  $U(I)$ . Middle: non-linear element with current-voltage characteristic  $I(U)$ . Right: exemplary current-voltage characteristic.

- **Impedance converter (DC-transformer):** If a DC-transformer with parameter  $1 : a$  has at one end the voltage-current pair  $(U_1, I_1)$  and at the other the voltage current pair  $(U_2, I_2)$ , then  $U_2 = aU_1$  and  $I_2 = -(1/a)I_1$  independent of time. Fig. 1.12(left) shows the element and Fig. 1.12(middle and right) the voltage-voltage and current-current characteristics. In the vector-case the relations are  $\mathbf{U}_2 = \mathbf{A}\mathbf{U}_1$  and  $\mathbf{I}_2 = -\mathbf{A}^{-T}\mathbf{I}_1$ . If the label  $1 : a$  is missing then the parameter is  $1 : 1$  and the slopes are

+1 and  $-1$ , respectively. The DC-transformer is equivalent to the impedance converter which is shown in Fig. 1.13. Its  $[ABCD]$  chain-matrix is<sup>5</sup>

$$[ABCD] = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \quad (1.14)$$

where (note that there is a minus in front of  $I_1$ )

$$\begin{pmatrix} U_2 \\ +I_2 \end{pmatrix} = [ABCD] \cdot \begin{pmatrix} U_1 \\ -I_1 \end{pmatrix} \quad (1.15)$$

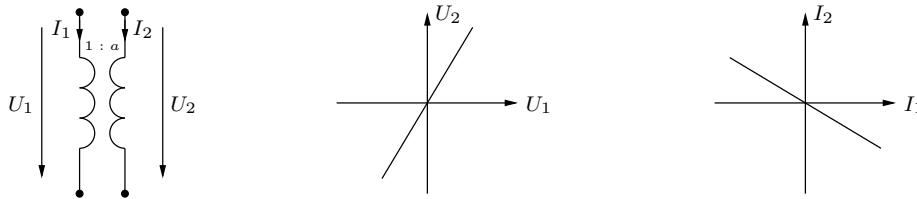


Figure 1.12: Left: DC-transformer with parameter  $1 : a$ . Middle: voltage-voltage characteristic (the slope is  $a$ ). Right: current-current characteristic (the slope is  $-1/a$ ). Note: if the label  $1 : a$  is missing, then the parameter is  $1 : 1$ .

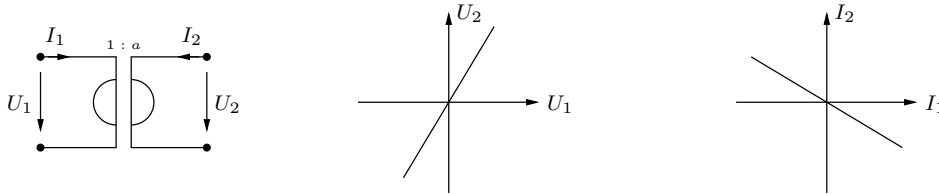


Figure 1.13: Left: Impedance converter with parameter  $1 : a$  (this element is equivalent to a DC-transformer with parameter  $1 : a$ ). Middle: voltage-voltage characteristic (the slope is  $a$ ). Right: current-current characteristic (the slope is  $-1/a$ ). Note: if the label  $1 : a$  is missing, then the parameter is  $1 : 1$ .

- **Impedance inverter (Gyrator):** A gyrator with parameter  $g$  has at one end the voltage-current pair  $(U_1, I_1)$  and at the other the voltage-current pair  $(U_2, I_2)$  and is described by the chain matrix

$$[ABCD] = \begin{pmatrix} 0 & 1/g \\ g & 0 \end{pmatrix}, \quad (1.16)$$

independent of time. Fig. 1.14(left) shows the element and Fig. 1.14(middle and right) the current(2)-voltage(1) and the voltage(2)-current(1) characteristics. In the vector-case the relations are  $\mathbf{I}_2 = \mathbf{G}\mathbf{U}_1$  and  $\mathbf{U}_2 = -\mathbf{G}^{-T}\mathbf{I}_1$ .

- **Capacitor:** A capacitor with parameter  $C$  is a time-dependent electrical element with current-voltage characteristic  $I(U(\cdot), t) = C \cdot \partial U(t)/\partial t$  and voltage-current characteristic  $U(I(\cdot), t) = (1/C) \cdot \int_{-\infty}^t I(t') dt'$ . Fig. 1.15(a) shows the element.

<sup>5</sup>The  $A, B, C, D$  of  $[ABCD]$  have nothing to do with the same letters used for the Kalman filter.

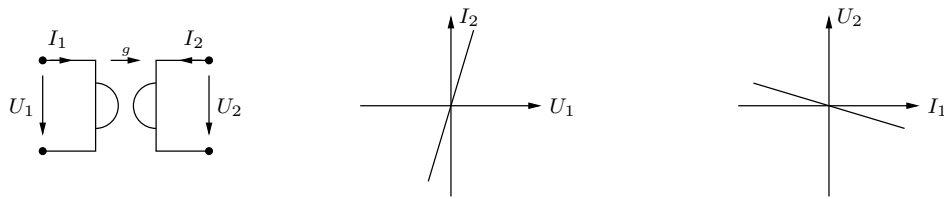


Figure 1.14: Left: Gyrator with parameter  $g$ . Middle: current(2)-voltage(1) characteristic (the slope is  $g$ ). Right: voltage(2)-current(1) characteristic (the slope is  $-1/g$ ).

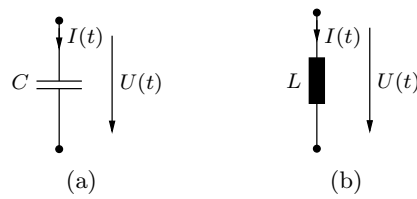


Figure 1.15: (a) Capacitor. (b) Inductor.

- **Inductor:** An inductor with parameter  $L$  is a time-dependent electrical element with voltage-current characteristic  $U(I(\cdot), t) = L \cdot \partial I(t) / \partial t$  and current-voltage characteristic  $I(U(\cdot), t) = (1/L) \cdot \int_{-\infty}^t U(t') dt'$ . Fig. 1.15(b) shows the element.

## Chapter 2

# Simple Motivating Examples

The aim of this section is to give some simple examples which should give the reader a flavor of the results that will follow in this report.

### 2.1 Example 1 (Summation Function Node)

Figure 2.1 shows a Forney-type factor graph (FFG) involving the random variables  $X_1$ ,  $X_2$  and  $Y$ . The global function represents the joint density of  $X_1$ ,  $X_2$  and  $Y$  and is the product of local functions:

$$p_{X_1 X_2 Y}(x_1, x_2, y) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot \delta(y - x_1 - x_2). \quad (2.1)$$

$\delta(\cdot)$  denotes the Dirac-delta distribution. Based on a measurement we would like to find the blockwise MAP estimate  $(\hat{x}_1, \hat{x}_2)$  of the vector  $(x_1, x_2)$ . (Note that we maximize jointly, see also the comments in Sec. 1.4.) To be specific, we assume  $X_1 \sim \mathcal{N}(0, \sigma_1^2)$  and  $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ :

$$p_{X_1 X_2 Y}(x_1, x_2, y) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-x_1^2/2\sigma_1^2) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-x_2^2/2\sigma_2^2) \cdot \delta(y - x_1 - x_2). \quad (2.2)$$

Given  $Y = y$ , maximizing  $p_{X_1 X_2 Y}(x_1, x_2, y)$  is equivalent to minimizing  $-\ln p_{X_1 X_2 Y}(x_1, x_2, y)$  or to the minimization of  $-\ln p_{X_1}(x_1) - \ln p_{X_2}(x_2)$  under the constraint  $y = x_1 + x_2$ . Using the technique of Lagrange multipliers we have to minimize

$$L := -\ln p_{X_1}(x_1) - \ln p_{X_2}(x_2) + \lambda(y - x_1 - x_2). \quad (2.3)$$

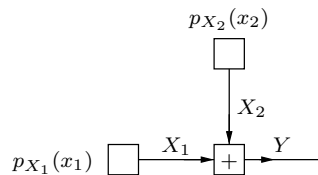


Figure 2.1: Example 1.

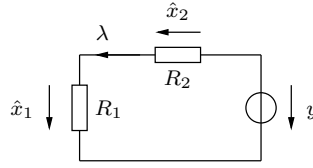


Figure 2.2: Electrical network for Example 1.

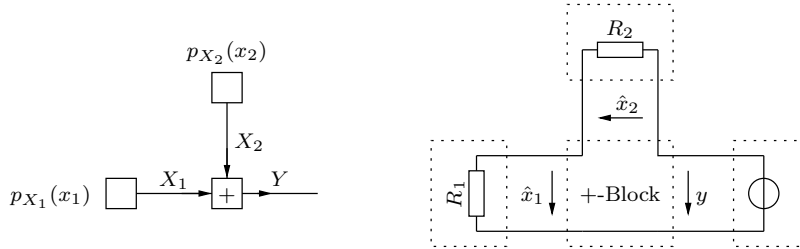


Figure 2.3: FFG versus electrical network in the case of Example 1.

In order to minimize  $L$ , we set the gradient equal to zero.

$$\begin{cases} \frac{\partial}{\partial x_1} L = \frac{x_1}{\sigma_1^2} - \lambda & \stackrel{!}{=} 0 & \text{(component law)} \\ \frac{\partial}{\partial x_2} L = \frac{x_2}{\sigma_2^2} - \lambda & \stackrel{!}{=} 0 & \text{(component law)} \\ \frac{\partial}{\partial \lambda} L = y - x_1 - x_2 & \stackrel{!}{=} 0 & \text{(Kirchhoff voltage law)} \end{cases} \quad (2.4)$$

The electrical network in Fig. 2.2 implements the three equations from Eq. (2.4) with  $R_1 = \sigma_1^2$  and  $R_2 = \sigma_2^2$ . The first two equations correspond to component equations, giving current-voltage characteristics of components, whereas the third equation describes a Kirchhoff voltage law (the sum of the voltages around a loop must be zero). In Fig. 2.3 the electrical network is redrawn so as to show its close topological relationship to its FFG. The Lagrange multiplier  $\lambda$  turns out to be a current and loosely speaking plays the role of “exchanging information” from one part of the circuit to the other.

Assume that we are only interested in the estimate  $\hat{x}_1$ . We can modify as long as we do not touch (i.e. we do not remove) the nodes between which we have the voltage  $\hat{x}_1$ . Fig. 2.4 shows possible modifications that can be performed. After these modifications we obtain a network

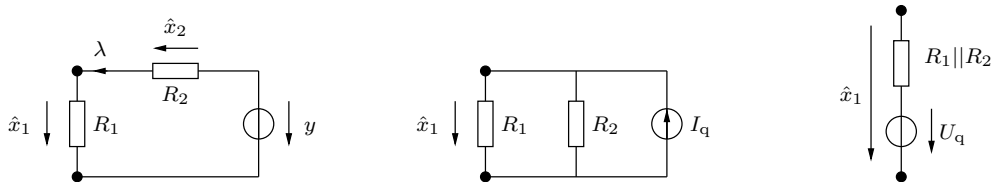


Figure 2.4: Modifications of the electrical network of Example 1 leaving the highlighted nodes untouched.  $(R_1 || R_2) = \frac{R_1 R_2}{R_1 + R_2}$ ,  $I_q = \frac{1}{R_2} y$ , and  $U_q = (R_1 || R_2) I_q = \frac{R_1}{R_1 + R_2} y$  have been introduced.

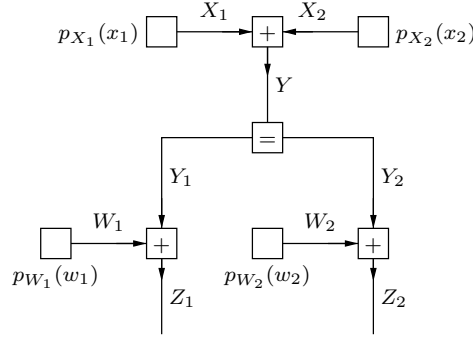


Figure 2.5: Example 2.

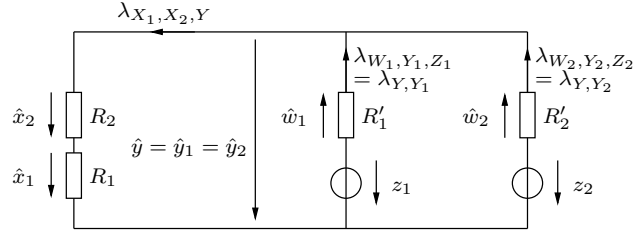


Figure 2.6: Example 2.

in which no current can flow anymore, so we can directly read off the estimate  $\hat{x}_1$ . In Ch. 4 we will see that these operations basically correspond to performing the MPA (see Sec. 1.4). In the case of jointly Gaussian random variables two special things happen: the MPA is equivalent to the SPA (see Sec. 1.3) and we can also directly read off the entries of the error covariance matrix (see Sec. 5.3).

## 2.2 Example 2 (Summation-Function and Equality-Function Nodes)

Figure 2.5 shows an FFG involving the random variables  $X_1$ ,  $X_2$ ,  $W_1$ ,  $W_2$ ,  $Y$ ,  $Y_1$ ,  $Y_2$ ,  $Z_1$ , and  $Z_2$ . The global function represents the joint density of these random variables and is the product of local functions:

$$\begin{aligned}
 & p_{X_1 X_2 W_1 W_2 Y Y_1 Y_2 Z_1 Z_2}(x_1, x_2, w_1, w_2, y, y_1, y_2, z_1, z_2) \\
 &= p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot p_{W_1}(w_1) \cdot p_{W_2}(w_2) \cdot \\
 & \quad \delta(y - x_1 - x_2) \cdot \delta(y_1 - y) \cdot \delta(y_2 - y) \cdot \delta(z_1 - y_1 - w_1) \cdot \delta(z_2 - y_2 - w_2). \quad (2.5)
 \end{aligned}$$

Note that the equality-function node is described by a product of two terms (generally, an equality-function node of degree  $d$  is described by a product of  $d - 1$  terms). Given the measurements  $Z_1 = z_1$  and  $Z_2 = z_2$  we want to make a blockwise MAP estimate of the vector  $(x_1, x_2, w_1, w_2, y, y_1, y_2)$ . But maximizing  $p(x_1, x_2, w_1, w_2, y, y_1, y_2, z_1, z_2)$  is equivalent

to minimizing the Lagrangian

$$\begin{aligned}
L := & -\ln p_{X_1}(x_1) - \ln p_{X_2}(x_2) - \ln p_{W_1}(w_1) - \ln p_{W_2}(w_2) \\
& + \lambda_{X_1, X_2, Y}(y - x_1 - x_2) + \lambda_{Y, Y_1}(y_1 - y) + \lambda_{Y, Y_2}(y_2 - y) \\
& + \lambda_{Y_1, W_1, Z_1}(z_1 - w_1 - y_1) + \lambda_{Y_2, W_2, Z_2}(z_2 - w_2 - y_2)
\end{aligned} \tag{2.6}$$

In order to minimize  $L$  we set the gradient equal to zero.

$$\left\{ \begin{array}{ll}
\frac{\partial}{\partial x_1} L = -\frac{p'_{X_1}(x_1)}{p_{X_1}(x_1)} - \lambda_{X_1, X_2, Y} & \stackrel{!}{=} 0 \text{ (component law)} \\
\frac{\partial}{\partial x_2} L = -\frac{p'_{X_2}(x_2)}{p_{X_2}(x_2)} - \lambda_{X_1, X_2, Y} & \stackrel{!}{=} 0 \text{ (component law)} \\
\frac{\partial}{\partial w_1} L = -\frac{p'_{W_1}(w_1)}{p_{W_1}(w_1)} - \lambda_{Y_1, W_1, Z_1} & \stackrel{!}{=} 0 \text{ (component law)} \\
\frac{\partial}{\partial w_2} L = -\frac{p'_{W_2}(w_2)}{p_{W_2}(w_2)} - \lambda_{Y_2, W_2, Z_2} & \stackrel{!}{=} 0 \text{ (component law)} \\
\frac{\partial}{\partial y} L = \lambda_{X_1, X_2, Y} - \lambda_{Y, Y_1} - \lambda_{Y, Y_2} & \stackrel{!}{=} 0 \text{ (Kirchhoff current law)} \\
\frac{\partial}{\partial y_1} L = \lambda_{Y, Y_1} - \lambda_{Y_1, W_1, Z_1} & \stackrel{!}{=} 0 \text{ (current equivalence)} \\
\frac{\partial}{\partial y_2} L = \lambda_{Y, Y_2} - \lambda_{Y_2, W_2, Z_2} & \stackrel{!}{=} 0 \text{ (current equivalence)} \\
\frac{\partial}{\partial \lambda_{X_1, X_2, Y}} L = y - x_1 - x_2 & \stackrel{!}{=} 0 \text{ (Kirchhoff voltage law)} \\
\frac{\partial}{\partial \lambda_{Y, Y_1}} L = y_1 - y & \stackrel{!}{=} 0 \text{ (Kirchhoff voltage law)} \\
\frac{\partial}{\partial \lambda_{Y, Y_2}} L = y_2 - y & \stackrel{!}{=} 0 \text{ (Kirchhoff voltage law)} \\
\frac{\partial}{\partial \lambda_{Y_1, W_1, Z_1}} L = z_1 - w_1 - y_1 & \stackrel{!}{=} 0 \text{ (Kirchhoff voltage law)} \\
\frac{\partial}{\partial \lambda_{Y_2, W_2, Z_2}} L = z_2 - w_2 - y_2 & \stackrel{!}{=} 0 \text{ (Kirchhoff voltage law)}
\end{array} \right. \tag{2.7}$$

Assuming  $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ ,  $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ ,  $W_1 \sim \mathcal{N}(0, \sigma_1'^2)$ , and  $W_2 \sim \mathcal{N}(0, \sigma_2'^2)$  and defining  $R_1 := \sigma_1^2$ ,  $R_2 := \sigma_2^2$ ,  $R_1' := \sigma_1'^2$ , and  $R_2' := \sigma_2'^2$ , the system of equations in Eq. (2.7) is equivalent to the electrical network in Fig. 2.6. The first four equations describe component current-voltage laws, the next one a Kirchhoff current law, the next two equivalences of certain currents, and the last five describe Kirchhoff voltage laws. Note again, that there is a tight topological relationship between the FFG and the electrical network.

As we did for the example in Sec. 2.1, we can modify the electrical network in this example so that we can directly read off a desired estimate. Assume that we are interested in the estimate  $\hat{x}_2$ . We can modify the network as we like under the condition that the two nodes between which  $\hat{x}_2$  is measured are left untouched: we can transform the real voltage sources  $(z_1, R_1')$  and  $(z_2, R_2')$  into real current sources, combine these two real current sources to one real current source and replace it again by a real voltage source, and so on. Again, this is nothing else than performing the max-product algorithm (MPA, see Ch. 4).

### 2.3 Example 3 (Non-Gaussian Random Variables)

Let us now consider a slightly modified version of Example 1 in Sec. 2.1.  $X_1$  is still Gaussian with  $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ , but  $X_2$  is now distributed according to<sup>1</sup>

$$p_{X_2}(x_2) = k_{21} \frac{1}{\cosh(k_{22}x_2)} = k_{21} \exp(-\ln(\cosh(k_{22}x_2))), \tag{2.8}$$

<sup>1</sup>Note that  $\ln(\cosh(k_{22}x_2))$  is a convex function in  $x_2$ .



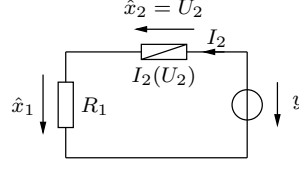


Figure 2.7: Example 3. The nonlinear resistor has characteristic  $I_2 = I_2(U_2) = k_{22} \tanh(k_{22}U_2)$ .

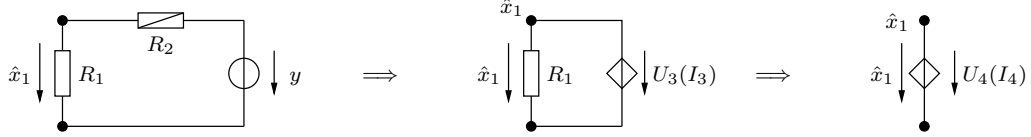


Figure 2.8: Example 3. The nonlinear resistor has characteristic  $I_2 = I_2(U_2) = k_{22} \tanh(k_{22}U_2)$ .

where  $k_{21}$  and  $k_{22}$  are chosen so that the mean is zero and the variance is  $\sigma_2^2$ . As before  $Y = X_1 + X_2$  and based on a measurement  $Y = y$  we want to get a blockwise MAP estimate  $(\hat{X}_1, \hat{X}_2)$  of  $(X_1, X_2)$ . The joint density of  $X_1$ ,  $X_2$ , and  $Y$  is

$$p_{X_1 X_2 Y}(x_1, x_2, y) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot \delta(y - x_1 - x_2) \quad (2.9)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-x_1^2/2\sigma_1^2\right) \cdot k_{22} \cosh(k_{22}(x_2)) \cdot \delta(y - x_1 - x_2) \quad (2.10)$$

$$= \frac{k_{22}}{\sqrt{2\pi}\sigma_1} \exp\left(-x_1^2/2\sigma_1^2 - \ln(\cosh(k_{22}x_2))\right) \cdot \delta(y - x_1 - x_2). \quad (2.11)$$

We proceed as in Example 1 of Sec. 2.1 and get the Lagrangian

$$L := -\ln p_{X_1}(x_1) - \ln p_{X_2}(x_2) + \lambda(y - x_1 - x_2), \quad (2.12)$$

In order to minimize  $L$  we set the gradient equal to zero.

$$\begin{cases} \frac{\partial}{\partial x_1} L = \frac{x_1}{\sigma_1^2} - \lambda & \stackrel{!}{=} 0 \quad (\text{component law}) \\ \frac{\partial}{\partial x_2} L = k_{22} \tanh(k_{22}(x_2)) - \lambda & \stackrel{!}{=} 0 \quad (\text{component law}) \\ \frac{\partial}{\partial \lambda} L = y - x_1 + x_2 & \stackrel{!}{=} 0 \quad (\text{Kirchhoff voltage law}) \end{cases} \quad (2.13)$$

which can be interpreted as the system of equation of the electrical network shown in Fig. 2.7. We would like to comment on different points.

- It is a convex optimization problem.
- It is not possible to solve algebraically for  $\hat{x}_1$  and  $\hat{x}_2$ .
- Resistor  $R_1$  is a “standard resistor” with current-voltage characteristic  $I_1 = I_1(U_1) = U_1/\sigma_1^2$ . But the current-voltage characteristic of the nonlinear resistor is

$$I_2 = I_2(U_2) = k_{22} \tanh(k_{22}U_2). \quad (2.14)$$

More about generalized resistors can be found in Secs. 5.1 and 5.4.

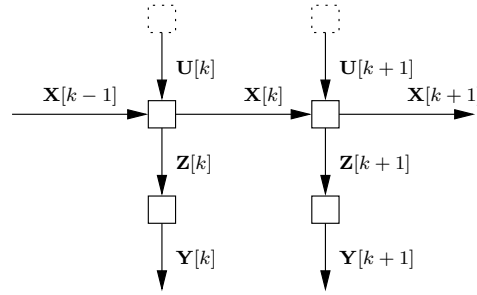


Figure 2.9: State space model.

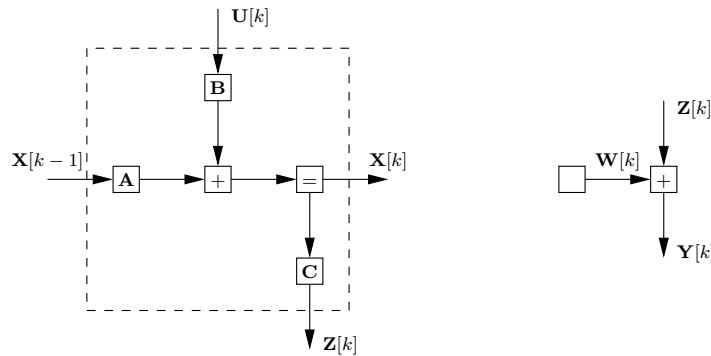


Figure 2.10: Details of linear state space model.

- This simple example shows that it is possible to get MAP estimates with the help of electrical networks also for problems where not all densities are Gaussian. But topologically the solution is the same as for Example 1 (see Fig. 2.3).
- Again, we can modify the circuit so that we can directly read off the solution. Figure 2.8 shows how this is done. Thereby we introduced the real voltage source with characteristic  $U_3(I) = y + \operatorname{artanh}(I_3/k_{22})/k_{22}$  and the real voltage source with characteristic  $U_4(I_4)$ .  $U_4(I_4)$  can easily be generated graphically by adding the  $U_3(I)$  and the  $U = R_1 I$  characteristics along the  $I$ -axis. The solution is then given by  $\hat{x}_1 = U_4(0)$ , because the current through in Fig. 2.8(right) must be zero. For more about this, see Ch. 4.

## 2.4 Example 4 (A Simple Kalman Filter)

In Kalman filtering (for a general reference, see e.g. [12]) we assume to have the following state-space model

$$\mathbf{x}[k] = \mathbf{A}[k]\mathbf{x}[k-1] + \mathbf{B}[k]\mathbf{u}[k] \quad (2.15)$$

$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{w}[k], \quad (2.16)$$

where  $u[k]$  and  $w[k]$  are stochastic processes. To get the intuition of the results, we mainly consider the scalar Kalman filter (SKF), meaning that  $u[k]$ ,  $w[k]$ ,  $x[k]$ ,  $y[k]$ ,  $a[k]$ ,  $b[k]$ , and

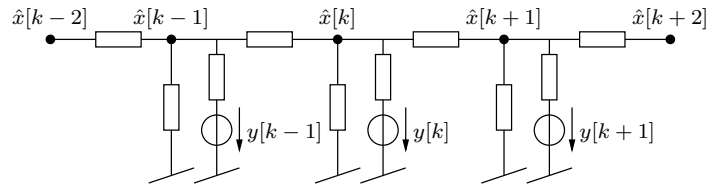


Figure 2.11: Several Sections of an electrical circuit implementing an SKF. The resistors have the values  $R_{k,k}$ ,  $R_{k,k-1}$ ,  $R_{k,k+1}$ ,  $R'_{k,k}$ , respectively.

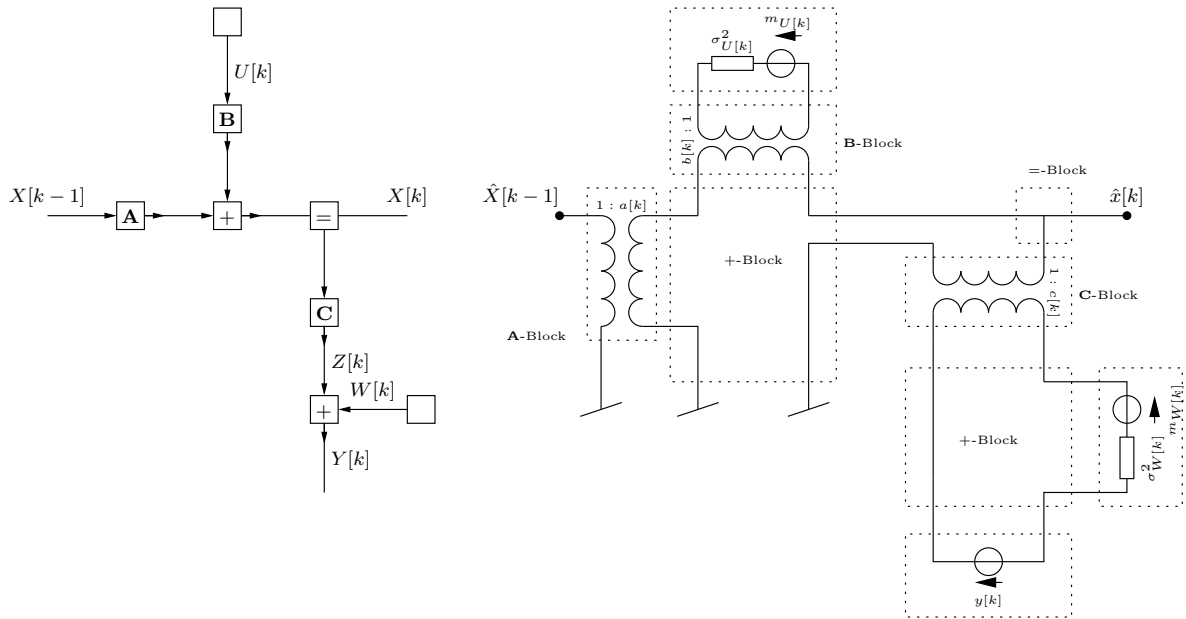


Figure 2.12: Left: FFG of one section of a Kalman filter. Right: Electrical network of a section of a Kalman filter highlighting the topological equivalence with the FFG.

$c[k]$  are all scalars. So we assume to have the state-space model

$$x[k] = a[k]x[k-1] + u[k] \quad (2.17)$$

$$y[k] = c[k]x[k] + w[k], \quad (2.18)$$

where we replaced without loss of generality  $b[k]u[k]$  by  $u[k]$ . Furthermore, let  $u[k] \sim \mathcal{N}(0, \sigma_{U[k]}^2)$  and  $w[k] \sim \mathcal{N}(0, \sigma_{W[k]}^2)$ . FFGs depicting a Kalman filter are shown in Figs. 2.9 and 2.10.

The task in Kalman filtering is to get an estimate  $\hat{x}[k]$  about  $x[k]$  based on a measurement of  $Y[k] = y[k]$  for  $N_1 \leq k \leq N_2$ . Let  $\mathbf{x}$  be the vector  $(x[N_1], \dots, x[N_2])$ ,  $\mathbf{y}$  be the vector  $(y[N_1], \dots, y[N_2])$ . (We do not worry about ‘‘border effects’’ in this report as we are mainly interested in the case  $N_1 \rightarrow -\infty$  and  $N_2 \rightarrow +\infty$ ). To get the blockwise<sup>2</sup> MAP estimate we calculate the joint density

$$p_{\mathbf{U}\mathbf{W}\mathbf{X}\mathbf{Y}}(\mathbf{u}, \mathbf{w}, \mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^N \prod_k \sigma_{U[k]} \prod_k \sigma_{W[k]}} \exp \left( - \sum_k \frac{1}{2\sigma_{U[k]}^2} u[k]^2 - \sum_k \frac{1}{2\sigma_{W[k]}^2} w[k]^2 \right) \cdot \delta(x[k] - a[k]x[k-1] - u[k]) \cdot \delta(y[k] - c[k]x[k] - w[k]). \quad (2.19)$$

This time we follow a slightly different approach. We recognize that maximizing  $p_{\mathbf{U}\mathbf{W}\mathbf{X}\mathbf{Y}}(\mathbf{u}, \mathbf{w}, \mathbf{x}, \mathbf{y})$  is equivalent to maximizing  $p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ , which is<sup>3</sup>

$$p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^N \prod_k \sigma_{U[k]} \prod_k \sigma_{W[k]}} \cdot \exp \left( - \sum_k \frac{1}{2\sigma_{U[k]}^2} (x[k] - a[k]x[k-1])^2 - \sum_k \frac{1}{2\sigma_{W[k]}^2} (y[k] - c[k]x[k])^2 \right). \quad (2.20)$$

Note that the exponent consists of local terms only, i.e. a summand which contains variables and constants of time step  $k$ , contains at most also variables and constants from time steps  $k-1$  and  $k+1$ . This is key to getting finally a simply wired electrical circuit. We solve

$$-\frac{\partial}{\partial \mathbf{x}} \ln p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \stackrel{!}{=} 0. \quad (2.21)$$

which amounts to the system of equations

$$\frac{x[k] - a[k]x[k-1]}{\sigma_{U[k]}^2} - \frac{(x[k+1] - a[k+1]x[k])a[k+1]}{\sigma_{U[k+1]}^2} - \frac{(y[k] - c[k]x[k])c[k]}{\sigma_{W[k]}^2} \stackrel{!}{=} 0, \quad (2.22)$$

for all  $k$ , or, equivalently

$$\left( \frac{1}{\sigma_{U[k]}^2} + \frac{a[k+1]^2}{\sigma_{U[k+1]}^2} + \frac{c[k]^2}{\sigma_{W[k]}^2} \right) x[k] - \frac{a[k]}{\sigma_{U[k]}^2} x[k-1] - \frac{a[k+1]}{\sigma_{U[k+1]}^2} x[k+1] - \frac{c[k]}{\sigma_{W[k]}^2} y[k] \stackrel{!}{=} 0, \quad (2.23)$$

<sup>2</sup>Blockwise means that we estimate the  $x[k]$  for all  $k$  of interest at the same time.

<sup>3</sup>This ‘‘elimination’’ of  $\mathbf{U}$  and  $\mathbf{W}$  does not change the solution  $\hat{\mathbf{x}}$  because the random vectors  $\mathbf{U}$  and  $\mathbf{W}$  can be determined from the remaining random vectors.

or,

$$\left( \frac{1 - a[k]}{\sigma_{U[k]}^2} + \frac{a[k+1]^2 - a[k+1]}{\sigma_{U[k+1]}^2} + \frac{c[k]^2 - c[k]}{\sigma_{W[k]}^2} \right) x[k] \quad (2.24)$$

$$+ \frac{a[k]}{\sigma_{U[k]}^2} (x[k] - x[k-1]) + \frac{a[k+1]}{\sigma_{U[k+1]}^2} (x[k] - x[k+1]) + \frac{c[k]}{\sigma_{W[k]}^2} (x[k] - y[k]) = 0, \quad (2.25)$$

which can be written as

$$R_{k,k}^{-1} x[k] + R_{k,k-1}^{-1} (x[k] - x[k-1]) + R_{k,k+1}^{-1} (x[k] - x[k+1]) + R'_{k,k}^{-1} (x[k] - y[k]) = 0. \quad (2.26)$$

But again, interpreting Eq. (2.26) as Kirchhoffs current law for node  $\hat{x}[k]$ , we obtain an electrical circuit as shown in Figure 2.11 where several sections of an SKF are shown.

Had we proceeded as in the example of Sec. 2.1, we would have got a circuit as in Fig. 2.12(right).

Using network transformation rules it is possible to convert the electrical network in Fig. 2.11 into the electrical network in Fig. 2.12(right) and vice-versa.

A more systematic approach to get the electrical network would be to write  $-\ln p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$  as

$$\frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} - \mathbf{x}^T \mathbf{j} + \text{const.}, \quad (2.27)$$

where  $\mathbf{G}$  has the diagonal entries

$$G_{k,k} = \frac{a[k+1]^2}{\sigma_{U[k+1]}^2} + \frac{1}{\sigma_{U[k]}^2} + \frac{c[k]^2}{\sigma_{W[k]}^2}, \quad (2.28)$$

and the non-diagonal entries

$$G_{k-1,k} = G_{k,k-1} = -\frac{a[k]}{\sigma_{U[k]}^2}. \quad (2.29)$$

All other entries of  $\mathbf{G}$  are zero. The vector  $\mathbf{j}$  has the entries

$$j_k = -\frac{a[k+1]m_{U[k+1]}}{\sigma_{U[k+1]}^2} + \frac{m_{U[k]}}{\sigma_{U[k]}^2} + \frac{c[k](y[k] - m_{W[k]})}{\sigma_{W[k]}^2}, \quad (2.30)$$

or

$$j_k = \frac{c[k]y[k]}{\sigma_{W[k]}^2} \quad (2.31)$$

in case all means are zero. Solving Eq. (2.21) gives the system of linear equations

$$\mathbf{G} \cdot \hat{\mathbf{x}} = \mathbf{j}. \quad (2.32)$$

Using the techniques from App. B we get the same electrical network as before (in fact, the steps we performed in Eqs. (2.22)–(2.26) are equivalent to the steps performed in App. B.



## Chapter 3

# From a Forney-style Factor Graph to an Electrical Network

After having studied some examples in Chapter 2, we give now the general derivation to derive an electrical network solving a blockwise MAP problem stated with the help of an FFG. The characteristics of the electrical networks used in this chapter are given in Sec. 1.7.

### 3.1 The Overall Lagrangian

We assume that we have an FFG which describes the joint density of the random variables  $\mathbf{X} = (X_1, \dots, X_n)^T$ . The vector  $\mathbf{X}$  consists of two parts  $\mathbf{X} = (\mathbf{X}^{\text{est}T}, \mathbf{X}^{\text{meas}T})^T$ :

- $\mathbf{X}^{\text{est}}$ : random variables that have to be estimated,
- $\mathbf{X}^{\text{meas}}$ : random variables that are measured.

Based on the measurement of  $\mathbf{X}^{\text{meas}} = \mathbf{x}^{\text{meas}}$  we want to get the (blockwise) MAP estimate  $\hat{\mathbf{X}}^{\text{est}}$  of  $\mathbf{X}^{\text{est}}$ . We assume that the FFG involves four types of function nodes: leaf-function nodes<sup>1</sup>, summation-function nodes, equality-function nodes, and scaling-function nodes (see Fig. 3.1)<sup>2</sup>.

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{t \in \mathcal{F}_1} f_t(x_t) \cdot \prod_{t \in \mathcal{F}_2} f_t(\mathbf{x}_t) \cdot \prod_{t \in \mathcal{F}_3} f_t(\mathbf{x}_t) \cdot \prod_{t \in \mathcal{F}_4} f_t(x_{t,1}, x_{t,2}), \quad (3.1)$$

<sup>1</sup>By a leaf-function node  $f_t(\cdot)$  we mean a function node of degree one. We assume that  $-\ln f_t(\cdot)$  is a convex function.

<sup>2</sup>Note that here the summation-function node is defined in a symmetric way.

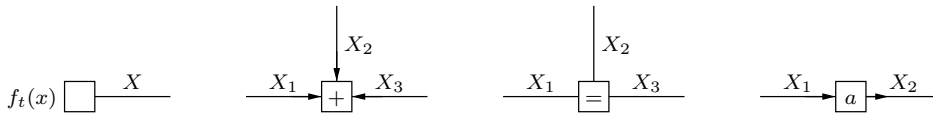


Figure 3.1: Different function nodes in an FFG: function node describing a distribution function, summation-function node (the only purpose of the arrows is to indicate the direction of the summation), equality-function node, and scaling-function node.

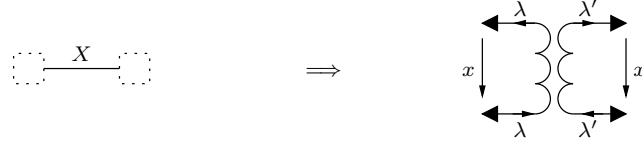


Figure 3.2: From FFG to electrical network: edge (case 1).

where  $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{F}_4$ , are the sets of indices of the one-degree function nodes, of summation-function nodes, of equality-function nodes, and of scaling-function nodes, respectively.  $\mathbf{x}_t$  denotes the arguments of the function  $f_t(\cdot)$ .

Given the measurement  $\mathbf{X}^{\text{meas}} = \mathbf{x}^{\text{meas}}$ , we want to maximize  $p_{\mathbf{X}}(\mathbf{x})$ , which is equivalent to minimizing  $-\ln p_{\mathbf{X}}(\mathbf{x})$ . But this is equivalent to minimizing the Lagrangian

$$\tilde{L} := - \sum_{t \in \mathcal{F}_1} \ln f_t(x_t) + \sum_{t \in \mathcal{F}_2} \lambda_t \left( \sum_s x_{t,s} \right) + \sum_{t \in \mathcal{F}_3} \sum_{s > 1} \lambda_{t,s} (x_{t,s} - x_{t,1}) \quad (3.2)$$

$$+ \sum_{t \in \mathcal{F}_4} \lambda_t (x_{t,2} - a_t x_{t,1}) + \sum_{e \in \mathcal{E}_5} \lambda_e (x_e - x_e^{\text{meas}}), \quad (3.3)$$

where  $\mathcal{E}_5$  denotes the set of edges whose value is measured and  $x_e^{\text{meas}}$  the value of the measured edge  $X_e$ . For deriving the general result it is convenient to symmetrize the expressions involving the equality-function nodes and the scaling function nodes: for each such node we introduce a help variable<sup>3</sup>  $x_t^{\text{h}}$  and obtain the new Lagrangian

$$L := \sum_{t \in \mathcal{F}_1} \ln f_t(x_t) + \sum_{t \in \mathcal{F}_2} \lambda_t \left( \sum_s x_{t,s} \right) + \sum_{t \in \mathcal{F}_3} \sum_s \lambda_{t,s} (x_{t,s} - x_t^{\text{h}}) \quad (3.4)$$

$$+ \sum_{t \in \mathcal{F}_4} \left( \lambda_{t,1} (x_{t,1} - x_t^{\text{h}}) + \lambda_{t,2} (x_{t,2} - a_t x_t^{\text{h}}) \right) + \sum_{e \in \mathcal{E}_5} \lambda_e (x_e - x_e^{\text{meas}}), \quad (3.5)$$

The Lagrangian has been symmetrized so that for the help and measurement variables the typical term are  $-\lambda x^{\text{h}}$  and  $-\lambda x^{\text{meas}}$ , respectively, whereas for the other variables it is  $+\lambda x$ . In order to minimize  $L$  we set the gradient of  $L$  equal to zero. This is done in the next sections.

## 3.2 Edges

### 3.2.1 Edges Between Two Constraint Function Nodes

Let  $X$  be the name of the edge variable that is connected to two function nodes. The Lagrangian looks like

$$L = \dots + \lambda x + \dots + \lambda' x + \dots \quad (3.6)$$

<sup>3</sup>Introducing a help variable in a blockwise MAP problem does not change the solution as long as the help variable is completely determined by the other random variables.



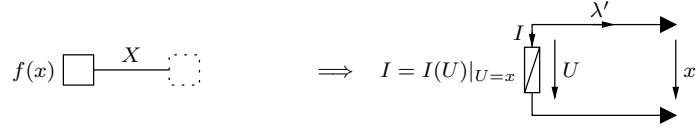


Figure 3.3: From FFG to electrical network: edge (case 2).

Deriving with respect to  $x$  and setting equal zero gives

$$\lambda + \lambda' \stackrel{!}{=} 0 \quad (3.7)$$

$$(3.8)$$

which can be interpreted as DC-transformer current law. The partial electrical network representing this equation is shown in Fig. 3.2. In general, a DC-transformer is necessary as the two sides may be on different potentials. Sometimes, DC-transformers can be removed (see also Sec. 3.7). Important is however that the current  $\lambda$  flowing out on the left top terminal flows in at the left bottom terminal; similarly for  $\lambda'$ . This will allow the multiport representation in Ch. 6.

### 3.2.2 Edges Between a Leaf-Function Node and a Constraint-Function Node

Let  $X$  be a variable edge that is connected to a one-degree function node  $f(\cdot)$  and a constraint function node. The Lagrangian looks like

$$L = \dots - \ln f(x) + \dots + \lambda'x + \dots \quad (3.9)$$

Deriving with respect to  $x$  and setting equal zero gives

$$-\frac{\partial}{\partial x} \ln(f(x)) + \lambda' \stackrel{!}{=} 0. \quad (3.10)$$

which can be interpreted as a component current-voltage characteristic

$$I = I(U) = -\frac{\partial}{\partial x} \ln f(x) \Big|_{x=U}. \quad (3.11)$$

The partial electrical network representing the above equation is shown in Fig. 3.3.

### 3.2.3 Edges Between Two One-Degree Function Nodes

This can be handled similarly to the case in Sec. 3.2.2.

## 3.3 Half-Edges

Let  $X$  be a variable edge that is connected to only one function node and assume that this variable is measured:  $x = x^{\text{meas}}$ . The Lagrangian looks like

$$L = \dots + \lambda(x - x^{\text{meas}}) + \dots + \lambda'x + \dots \quad (3.12)$$

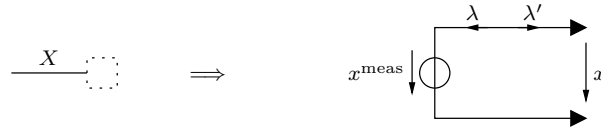


Figure 3.4: From FFG to electrical network: half-edge.

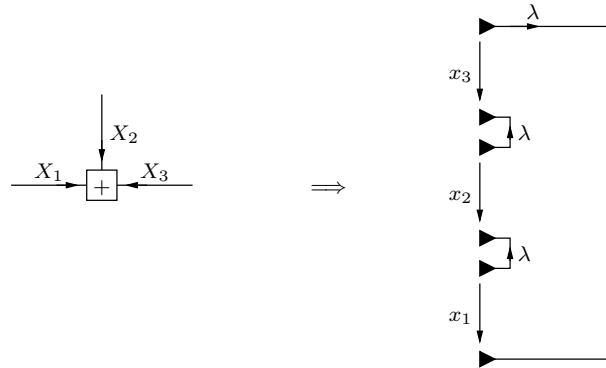


Figure 3.5: From FFG to electrical network: summation-function node.

Deriving  $L$  with respect to  $\lambda$  and  $x$ , respectively, and setting equal zero gives

$$\begin{cases} x - x^{\text{meas}} & \stackrel{!}{=} 0 \\ \lambda + \lambda' & \stackrel{!}{=} 0 \end{cases} \quad (3.13)$$

which can be interpreted as Kirchhoff voltage law and Kirchhoff current law, respectively. The partial electrical network representing this equation is shown in Fig. 3.4.

If  $X$  is *not measured*, then one gets the same partial electric network as in Fig. 3.4 with the voltage source *removed*.

### 3.4 Summation-Function Node

Let  $X_1, X_2, \dots, X_n$  be the variable edges attached to a summation-function node. To see the general principle, it is sufficient to consider the case  $n = 3$ . Then the Lagrangian looks like

$$L = \dots + \lambda(x_1 + x_2 + x_3) + \dots \quad (3.14)$$

Deriving  $L$  with respect to  $\lambda$  and setting equal zero gives

$$x_1 + x_2 + x_3 \stackrel{!}{=} 0, \quad (3.15)$$

which can be interpreted as Kirchhoff voltage law. The partial electrical network representing this equation is shown in Fig. 3.5. Note that  $\lambda$  is the current going out at one branch of a port and coming in at the other branch of the same port.<sup>4</sup>

<sup>4</sup>All ports of a summation function node are connected to partial electrical networks representing edges and half-edges. The above statement then follows from the properties of the partial electrical networks that replace edges and half-edges, see Secs. 3.2 and 3.3.

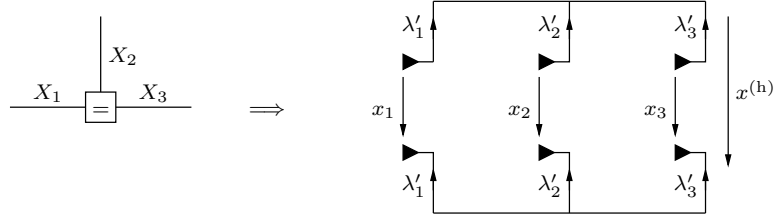


Figure 3.6: From FFG to electrical network: equality-function node.

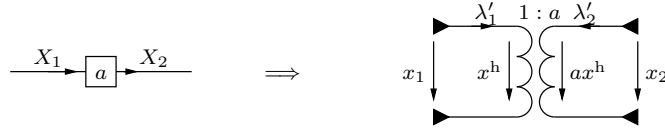


Figure 3.7: From FFG to electrical network: scaling-function node.

### 3.5 Equality-Function Node

Let  $X_1, X_2, \dots, X_n$  be the variable edges attached to an equality-function node. To see the general principle, it is sufficient to consider the case  $n = 3$ . Then the Lagrangian looks like

$$L = \dots + \lambda'_1(x_1 - x^h) + \lambda'_2(x_2 - x^h) + \lambda'_3(x_3 - x^h) + \dots \quad (3.16)$$

Deriving  $L$  with respect to  $\lambda'_1, \lambda'_2, \lambda'_3$ , and  $x^h$  respectively, and setting equal zero gives

$$\begin{cases} x_1 - x^h & \stackrel{!}{=} 0 \\ x_2 - x^h & \stackrel{!}{=} 0 \\ x_3 - x^h & \stackrel{!}{=} 0 \\ -\lambda'_1 - \lambda'_2 - \lambda'_3 & \stackrel{!}{=} 0, \end{cases} \quad (3.17)$$

where the first three equations can be interpreted as Kirchhoff voltage law and the last one as Kirchhoff current law. The partial electrical network representing this set of equations is shown in Fig. 3.6.

### 3.6 Scaling-Function Node

Let  $X_1$  and  $X_2$  be the variable edges attached to the scaling-function node which enforces  $X_2 = aX_1$ . The Lagrangian looks like

$$L = \dots + \lambda'_1(x_1 - x^h) + \lambda'_2(x_2 - ax^h) + \dots \quad (3.18)$$

Deriving  $L$  with respect to  $\lambda'_1, \lambda'_2$ , and  $x^h$ , respectively, and setting equal zero gives

$$\begin{cases} x_1 - x^h & \stackrel{!}{=} 0 \\ x_2 - ax^h & \stackrel{!}{=} 0 \\ -\lambda'_1 - a\lambda'_2 & \stackrel{!}{=} 0 \end{cases} \quad (3.19)$$

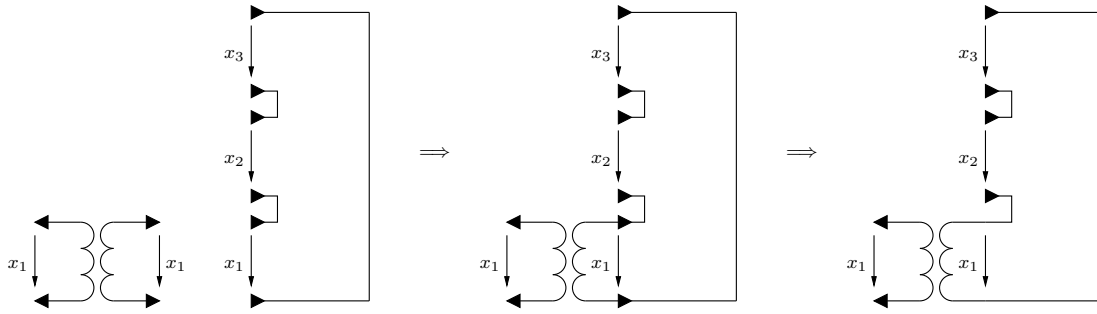


Figure 3.8: From FFG to electrical network: stitching two partial electrical networks together.

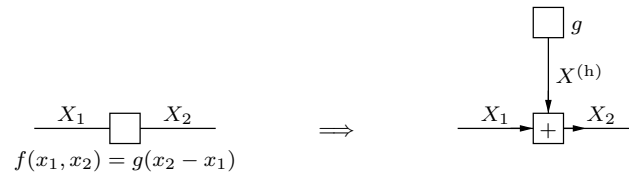


Figure 3.9: If a degree-two function node has a special form, i.e.,  $f(x_1, x_2) = g(x_2 - x_1)$  for some  $g(\cdot)$  then this FFG replacement rules gives an FFG using elements that have been treated in this report.

where the first two equations can be interpreted as Kirchhoff voltage law and the last one as DC-transformer current law. The partial electrical network representing this set of equations is shown in Fig. 3.7.

### 3.7 Putting It All Together

In the preceding sections we have derived the Lagrangian  $L$  with respect to all the different variables and derived various partial electrical networks. To get the overall electrical network one simple has to stitch these partial solutions together such that the filled triangles match. Then the filled triangles can be taken away as shown in Fig. 3.8 for a partial electrical network for an edge and a partial electrical network for a summation node. It is not difficult to see that the currents are the same on the wires that were joined together. We conclude with the following remarks:

- One can try to eliminate the 1 : 1 DC-transformers. (One is usually able to take away quite a few.)
- One can also try to eliminate the DC-transformers that are not 1 : 1 (see e.g. Sec. 2.4).
- The derived electrical networks are not unique in the following sense: if one eliminates random variables that are determined by the remaining random variables (see e.g. Sec. 2.4), then different electrical networks will result.
- The unit of the Lagrangian itself is power (or voltage squared divided by a resistor, or voltage times current). We have chosen the estimated and measured random variables

to be voltages. In this case the Lagrange multipliers turn out to be currents. Very loosely speaking, they give the amount that must be “exchanged” between different parts of an electrical network so that it is in a stationary condition.

- Note that the electrical networks derived in this chapter always give the correct block-wise MAP solution, independent of the fact if the underlying FFG *is a tree or not*. Important however is that the problem is convex and there exists only one optimum. More about this can be found in Chs. 4 and 7) and e.g. in [13].
- It is possible to allow functions of the form  $f(x_1, x_2) = g(x_2 - x_1)$  or  $f(x_1, x_2) = g(x_2 - ax_1)$ . In the first case, the idea is to introduce the help variable  $x^{(h)} = x_2 - x_1$  and replace the function node  $f(x_1, x_2)$  by a summation node  $\delta(x_2 - x_1 - x^{(h)})$ , a leaf function node  $g(x^{(h)})$  and the edge  $X^{(h)}$  inbetween, see Fig. 3.9. Essentially, this type of function node will result in a (generally nonlinear) element in the electrical network. The second case is treated similarly, using in addition a DC-transformer.
- Function nodes as discussed in the previous point occur especially in FFGs which are not trees. The overall function of the FFG is usually not a probability density or pmf anymore, but simply a non-negative function that could be turned into a probability density or pmf using proper normalization.



## Chapter 4

# The Max-Product Algorithm and the Sum-Product Algorithm

As mentioned in Sec. 1.4, the max-product algorithm (MPA) can be used to solve blockwise MAP problems in an exact way in the cases where the FFG is a tree. In the case of loopy FFGs, this is not anymore the case (but Freeman and Weiss give cases where the estimate is correct [13]). Similar things can be said about the sum-product algorithm (SPA). As already mentioned in Sec. 1.4, the case of jointly Gaussian random variables is very special: the blockwise MAP estimation problem is equivalent to the symbolwise MAP estimation problem, and the MPA and SPA are equivalent (see also [8]).

In the previous chapter we started with an FFG and derived an electrical network which solves the blockwise MAP estimation problem of the FFG. The purpose of this chapter is to show that network simplification (when the underlying FFG is a tree) corresponds one-to-one to the MPA (we saw two examples in Fig. 2.4 of Example 1 in Sec. 2.1 and in Fig. 2.8 of Example 3 in Sec. 2.3).

Note however that the electrical networks given in Ch. 3 always give the correct blockwise MAP solution, independent of the fact if the underlying FFG is a tree or not. Important however is that the problem is convex and there exists only one optimum (see also Ch. 7).

We have already seen in Ch. 3, that the negative derivative of the exponent of a density is an important function. We will see that it is a very handy function when performing the MPA. We therefore introduce the score function  $s(x)$  of a message  $\mu(x)$

$$s(x) = -\frac{\partial}{\partial x} \ln(\mu(x)). \quad (4.1)$$

From the score function  $s(x)$  we can get back to the message  $\mu(x)$  by the formula

$$\mu(x) \propto \exp\left(-\int_0^x s(x') dx'\right). \quad (4.2)$$

The proportionality constant (stemming basically from the integration constant) can be chosen so that  $\mu(x)$  fulfills a certain normalization, e.g.  $\int_{-\infty}^{+\infty} \mu(x) dx = 1$ . More about score functions can be found e.g. in [14].<sup>1</sup>, see also Sec. 5.1. An important special case is a Gaussian distributed message  $\mu(x) \propto \exp(-(x - m)^2/2\sigma^2)$ , whose score function and the inverse

---

<sup>1</sup>Note that in the book by Cover and Thomas the score function is defined as the positive derivative of the exponent of a density.

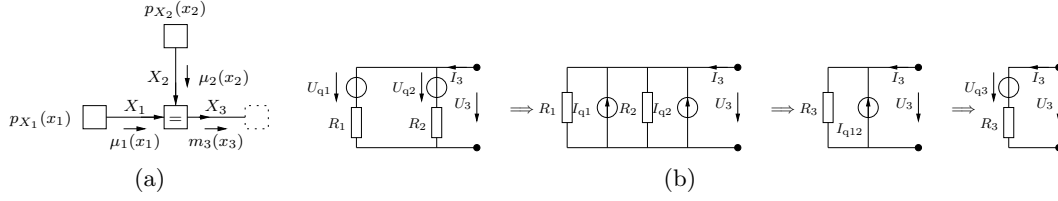


Figure 4.1: (a) Part of an FFG (the dotted box indicates the rest of the FFG). (b) Partial electrical network. Simplifying this part of the electrical network corresponds to calculating the message  $\mu_3(x_3)$ . We used the definitions  $I_{q1} := U_{q1}/R_1$ ,  $I_{q2} := U_{q1}/R_2$ ,  $R_3 := (R_1 || R_2) = R_1 R_2 / (R_1 + R_2)$ ,  $I_{q12} := I_{q1} + I_{q2}$ .

score functions are, respectively,

$$s(x) = \frac{(x - m)}{\sigma^2} = \frac{1}{\sigma^2}x - \frac{m}{\sigma^2}, \quad (4.3)$$

$$s^{-1}(y) = \sigma^2 y + m. \quad (4.4)$$

In this case, the score function graph is a straight line<sup>2</sup> with slope  $1/\sigma^2$ , i.e. the more confident the message is (the smaller  $\sigma^2$  is), the steeper is the score function.

In this chapter we will assume that the negative exponents of the messages are strictly convex, meaning that the score function is strictly monotonically increasing; this simplifies the proofs. But assuming only convexity would be enough to prove the same statements.

## 4.1 Examples

### 4.1.1 Equality-Function Node (Example)

We start with the simple example given in Fig. 4.1(a). We assume  $p_{X_1}(x_1)$  and  $p_{X_2}(x_2)$  are Gaussian distributions, so that the partial electrical network corresponding to them can be represented by a real voltage source (i.e. an ideal voltage source and a linear resistor in series), see Fig. 4.1(b). As we will see in this chapter, simplifying this part of the electrical network by following the steps indicated in Fig. 4.1 is then equivalent to calculating the message  $\mu_3(x_3)$ .

### 4.1.2 Summation-Function Node (Example)

This example is akin to the example given in Sec. 4.1.1, in the part of the FFG under consideration the equality-node function has been exchanged by a summation-function node (see Fig. 4.2(a)). This is yet another instance of the general fact that we will see later on, namely that calculating the message  $\mu_3(x_3)$  in the MPA corresponds to the electrical network simplifications shown in Fig. 4.2(b).

## 4.2 The General Case

We assume that the underlying FFG is a tree and that we are interested in the estimate of a specific variable  $X$ . Then the MPA starts sending messages at the the leaf function nodes

<sup>2</sup>This fact will be used in Sec. 4.3.



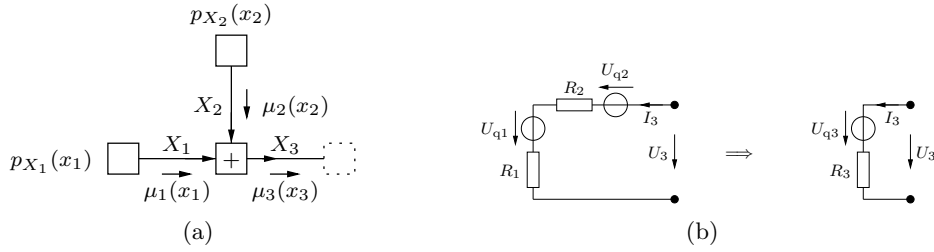


Figure 4.2: Same as Fig. 4.1, but now a summation-function node is under consideration. We used the definitions  $U_{q3} := U_{q1} + U_{q2}$ ,  $R_3 = R_1 + R_2$ .

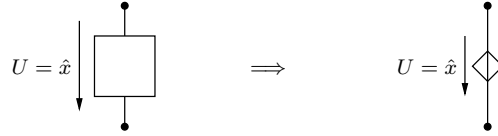


Figure 4.3: Left: the desired estimate is the voltage across a (possibly) complicated electrical network. Right: after the electrical network simplifications, which correspond to the MPA. Electrically, the two-terminal is not changed in the whole process.

and processes the messages in the interior according to the MPA rules until the edge  $X$  has received messages from both sides.

The aim of this section is to show that if the blockwise MAP problem is translated into an electrical network, the message passing necessary to get the estimate for  $X$  corresponds to simplifying the electrical network until the network consists only of a simple electrical element between the two nodes where the voltage corresponding to the estimate of  $X$  is measured. This is depicted in Fig. 4.3(left) where the desired estimate can be measured between the two nodes; inbetween is a (possibly) complicated electrical network. Simplifying the network between the two nodes the gives the electrical network in Fig. 4.3(right). Note that in the whole process we did not change the two nodes related to  $X$ , so as a two-terminal it behaves the same before and after the simplification. In the simplified electrical network it is trivial to read off the estimate from the remaining component between the two nodes, but nature gives us the possibility to read off the voltage already across the unsimplified network.

The following subsections discuss step by step how the electrical network simplification process works, and they show its relation to the MPA.

#### 4.2.1 Leaf Function Node (The General Case)

Fig. 4.4(left) shows the FFG of a leaf function node  $f_1(x_1)$ . According to the MPA, the message  $\mu_1(x_1)$  out of a leaf function node is the function itself  $f_1(x_1)$  itself. The score function  $s_1(x_1)$  of  $\mu_1(x_1)$  is calculated according to Eq. (4.1).

Fig. 4.4(right) shows the corresponding electrical network. We have seen in Sec. 3.2.2 that the current-voltage characteristic of a network element corresponds to the negative derivative of the logarithm of the exponent of the function  $f_1(\cdot)$ , i.e.  $I_1(U_1) = s_1(U_1)$ , where  $s_1(\cdot)$  is the score function corresponding to  $f_1(\cdot)$  as in the paragraph above. So the score function of the message and the current-voltage characteristic coincide and this will remain like this in the

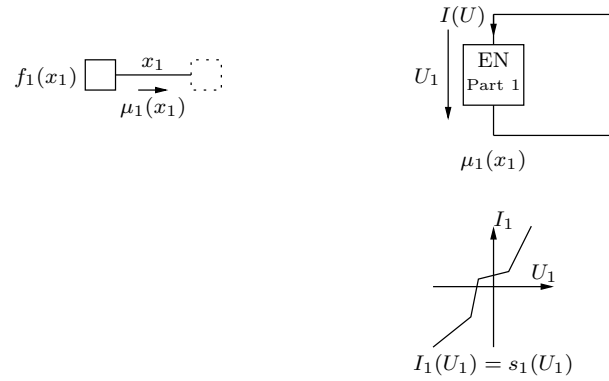


Figure 4.4: Leaf function node  $f_1(x_1)$ : FFG and electrical network.

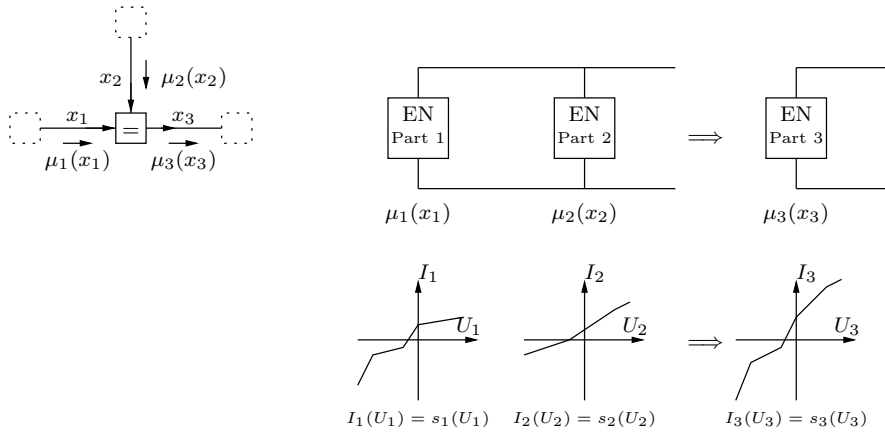


Figure 4.5: Left: MPA message-passing in the case of an equality-function node. Right: Equivalent simplification of parallelly connected partial electrical networks.

following steps.

As there is a one-to-one correspondence between messages and their score functions we can go back and forth between the two descriptions. In this case it is simple but important to note that the current-voltage characteristic describes (similar to the message) all we need to now about this part of the electrical network.

## 4.2.2 Equality-Function Node (The General Case)

Consider the FFG shown in Fig. 4.5(left). From the messages  $\mu_1(x_1)$  and  $\mu_2(x_2)$  we would like to calculate the message  $\mu_3(x_3)$  according to the MPA. By definition of the MPA, we

get<sup>3</sup>

$$\mu_3(x_3) : \propto \max_{x_1, x_2} f(x_1, x_2, x_3) \cdot \mu_1(x_1) \cdot \mu_2(x_2) \quad (4.5)$$

$$= \max_{x_1, x_2} \delta(x_3 - x_1) \cdot \delta(x_3 - x_2) \cdot \mu_1(x_1) \cdot \mu_2(x_2) \quad (4.6)$$

$$\propto \max_{x_1, x_2} [x_3 - x_1 = 0] \cdot [x_3 - x_2 = 0] \cdot \mu_1(x_1) \cdot \mu_2(x_2) \quad (4.7)$$

$$= \mu_1(x_3) \cdot \mu_2(x_3). \quad (4.8)$$

Using the score functions (see Eq. 4.1) we obtain (letting  $k$  be the appropriate constant)

$$s_3(x_3) := -\frac{\partial}{\partial x_3} \ln \mu_3(x_3) \quad (4.9)$$

$$= -\frac{\partial}{\partial x_3} \ln (k \cdot \mu_1(x_3) \cdot \mu_2(x_3)) \quad (4.10)$$

$$= -\frac{\partial}{\partial x_3} \ln (\mu_2(x_3)) - \frac{\partial}{\partial x_3} \ln (\mu_1(x_3)) \quad (4.11)$$

$$= s_1(x_3) + s_2(x_3), \quad (4.12)$$

which is a simple addition of the score functions.

We switch the perspective now and look at the the electrical network equivalent in Fig. 4.5(right) of the FFG in Fig. 4.5(left). In the way we started in Sec. 4.2.1 and by induction in Secs. 4.2.2, 4.2.3, and 4.2.4, we see that the score function  $s_1(x_1)$  of the message  $\mu_1(\cdot)$  corresponds to the current-voltage characteristic of the “electrical network part 1” and the score function  $s_2(x_2)$  of the message  $\mu_2(\cdot)$  corresponds to the current-voltage characteristic of the “electrical network part 2”. But combining two parallel electrical network parts whose current-voltage characteristic is known is very simple, namely

$$I_3(U_3) = I_1(U_3) + I_2(U_3). \quad (4.13)$$

(Eq. (4.13) is obviously equivalent to Eq. (4.12).) Therefore, we can replace the two parallel electrical networks by one electrical network with current-voltage characteristic  $I_3(U_3)$ .

### 4.2.3 Summation-Function Node (The General Case)

Consider the FFG shown in Fig. 4.6(left). From the messages  $\mu_1(x_1)$  and  $\mu_2(x_2)$  we would like to calculate the message  $\mu_3(x_3)$  according to the MPA. By definition of the MPA, we get

$$\mu_3(x_3) : \propto \max_{x_1, x_2} f(x_1, x_2, x_3) \cdot \mu_1(x_1) \cdot \mu_2(x_2) \quad (4.14)$$

$$= \max_{x_1, x_2} \delta(x_3 - x_1 - x_2) \cdot \mu_1(x_1) \cdot \mu_2(x_2) \quad (4.15)$$

$$\propto \max_{x_1, x_2} [x_3 - x_1 - x_2 = 0] \cdot \mu_1(x_1) \cdot \mu_2(x_2) \quad (4.16)$$

$$= \max_{x_1} \mu_1(x_1) \cdot \mu_2(x_3 - x_1). \quad (4.17)$$

This is obviously equivalent to

$$-\ln \mu_3(x_3) = -\ln \left( \max_{x_1} (\mu_1(x_1) \cdot \mu_2(x_3 - x_1)) \right) = \min_{x_1} \left( -\ln \mu_1(x_1) - \ln \mu_2(x_3 - x_1) \right). \quad (4.18)$$

---

<sup>3</sup>When  $S$  is a statement, then  $[S]$  equals 1 if the statement is true, otherwise  $[S]$  equals 0.

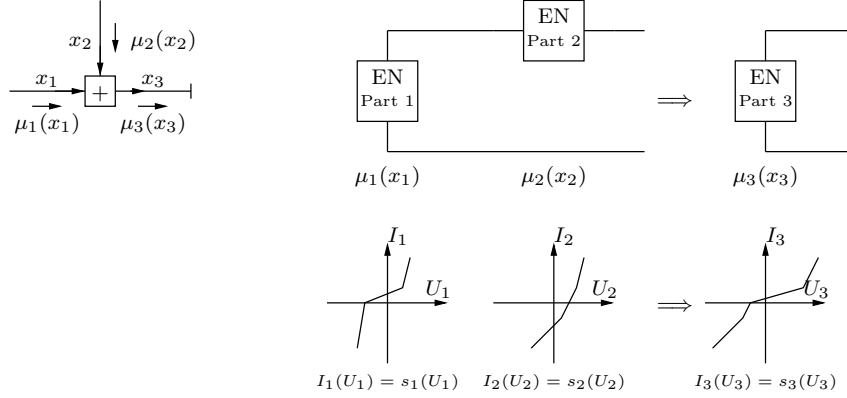


Figure 4.6: Left: MPA message-passing in the case of a summation-function node. Right: Equivalent simplification of serially connected partial electrical networks.

Assuming that both  $-\ln \mu_1(\cdot)$  and  $-\ln \mu_2(\cdot)$  are (strictly) convex functions, i.e., that both  $s_1(\cdot)$  and  $s_2(\cdot)$  are (strictly) increasing functions, we introduce the function<sup>4</sup>

$$x_1^*(x_3) := \arg \min_{x_1} (-\ln \mu_1(x_1) - \ln \mu_2(x_3 - x_1)), \quad (4.19)$$

which gives the  $x_1$  that minimizes  $-\ln \mu_1(x_1) - \ln \mu_2(x_3 - x_1)$  for a given  $x_3$ .<sup>5</sup> On the other hand, to find the minimizing  $x_1$  of  $-\ln \mu_1(x_1) - \ln \mu_2(x_3 - x_1)$  for a given  $x_3$ , we derive with respect to  $x_1$  and set equal to zero.

$$0 \stackrel{!}{=} \frac{\partial}{\partial x_1} (-\ln \mu_1(x_1) - \ln \mu_2(x_3 - x_1)) = s_1(x_1) - s_2(x_3 - x_1). \quad (4.20)$$

Combining Eqs. (4.19), and (4.20) we have

$$s_1(x_1^*(x_3)) = s_2(x_3 - x_1^*(x_3)), \quad (4.21)$$

and combining Eqs. (4.18) and (4.19) we get

$$-\ln \mu_3(x_3) = -\ln \mu_1(x_1^*(x_3)) - \ln \mu_2(x_3 - x_1^*(x_3)). \quad (4.22)$$

To find  $\mu_3(\cdot)$  and  $s_3(\cdot)$  we differentiate Eq. (4.22) with respect to  $x_3$ :

$$s_3(x_3) = \frac{\partial}{\partial x_3} (-\ln \mu_3(x_3)) \quad (4.23)$$

$$= \frac{\partial}{\partial x_3} (-\ln \mu_1(x_1^*(x_3)) - \ln \mu_2(x_3 - x_1^*(x_3))) \quad (4.24)$$

$$= s_1(x_1^*(x_3)) \cdot x_1^{*\prime}(x_3) + s_2(x_3 - x_1^*(x_3)) \cdot (1 - x_1^{*\prime}(x_3)) \quad (4.25)$$

$$= s_2(x_3 - x_1^*(x_3)). \quad (4.26)$$

<sup>4</sup>The notation  $x_1^*(\cdot)$  has nothing to do with Lagrange duality as in App. C.

<sup>5</sup>See the comment about convexity and strict convexity in the introductory part to this chapter. If we have “only” convexity, then  $\arg \min_{x_1}$  would give back a set.

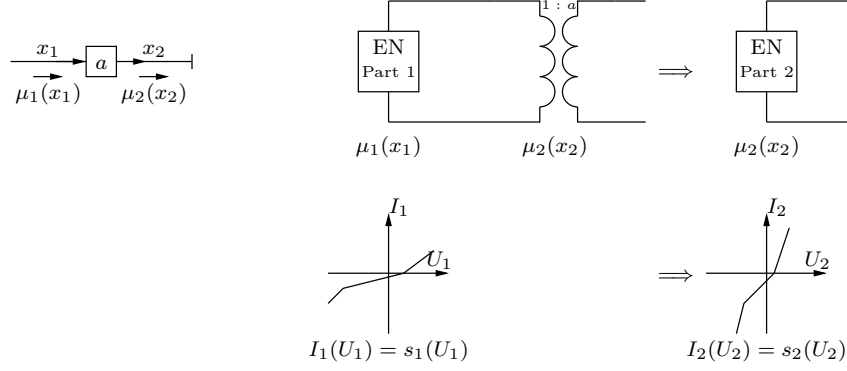


Figure 4.7: Left: MPA message-passing in the case of a scaling-function node. Right: Equivalent simplification of a partial electrical network and a DC-transformer ( $a = 0.5$  here).

With Eqs. (4.21) and (4.26) we conclude that

$$s_1(x_1^*(x_3)) = s_2(x_3 - x_1^*(x_3)) = s_3(x_3). \quad (4.27)$$

Going back to electrical networks we see that the electrical network equivalent of the FFG in Fig. 4.6(left) is the one in Fig. 4.6(right). In the way we started in Sec. 4.2.1, and by induction in Secs. 4.2.2, 4.2.3, and 4.2.4, we see that the score function  $s_1(x_1)$  of the message  $\mu_1(\cdot)$  corresponds to the current-voltage characteristic of the “electrical network part 1” and the score function  $s_2(x_2)$  of the message  $\mu_2(\cdot)$  corresponds to the current-voltage characteristic of the “electrical network part 2”. But combining two serially connected electrical network parts whose voltage-current characteristic<sup>6</sup> is known is very simple, namely

$$U_3(I_3) = U_1(I_3) + U_2(I_3), \quad (4.28)$$

(Eq. (4.28) is obviously equivalent to Eq. (4.27), but was derived with much more ease.) Therefore, we can replace the two serial electrical networks by one electrical network with current-voltage characteristic  $U_3(I_3)$ .

#### 4.2.4 Scaling-Function Node (The General Case)

Consider the FFG shown in Fig. 4.7(left). From the message  $\mu_1(x_1)$  we would like to calculate the message  $\mu_2(x_2)$  according to the MPA. By definition of the MPA, we get

$$\mu_2(x_2) \propto \max_{x_1} f(x_1, x_2) \cdot \mu_1(x_1) \quad (4.29)$$

$$= \max_{x_1} \delta(x_2 - ax_1) \cdot \mu_1(x_1) \quad (4.30)$$

$$\propto \max_{x_1} [x_2 - ax_1 = 0] \cdot \mu_1(x_1) \quad (4.31)$$

$$= \mu_1(x_2/a), \quad (4.32)$$

and therefore,

$$s_2(x_2) = -\frac{\partial}{\partial x_2} (\ln \mu_2(x_2)) = -\frac{\partial}{\partial x_2} (\ln \mu_1(x_2/a)) = \frac{1}{a} \cdot s_1(x_2/a). \quad (4.33)$$

<sup>6</sup>Note that inverting the current-voltage characteristic  $I(U)$  gives the voltage-current characteristic  $U(I)$ .

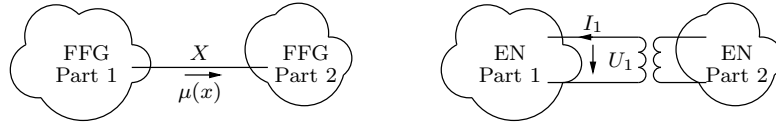


Figure 4.8: Left: In a FFG that is a tree, cutting any edge separates the FFG into two parts. The message  $\mu(\cdot)$  (or equivalently the score function) represents what part 2 has to know about part 1. Right: In an electrical network derived from an FFG which is a tree, the current-voltage characteristic  $I_1(U_1)$  represents what part 2 has to know about part 1.

In the context of electrical networks we see that the electrical network equivalent of the FFG in Fig. 4.7(left) is the one in Fig. 4.7(right). In the way we started in Sec. 4.2.1, and by induction in Secs. 4.2.2, 4.2.3, and 4.2.4, we see that the score function  $s_1(x_1)$  of the message  $\mu_1(\cdot)$  corresponds to the current-voltage characteristic of the “electrical network part 1”. But combining the partial electrical network, whose current-voltage characteristic is known, and the DC-transformer is very simple, namely

$$I_2(U_2) = \frac{1}{a} \cdot I_1(U_2/a), \quad (4.34)$$

(Eq. (4.34) is obviously the same as Eq. (4.33).) Therefore, we can replace the a partial electrical networks and a DC-transformer by one electrical network with current-voltage characteristic  $I_2(U_2)$ .

#### 4.2.5 Concluding Remarks (The General Case)

The update rules presented in the previous subsections clearly show the advantage of the score function associated to a message. The update rules are very simple geometrically:

- At an equality-function node the score functions have to be added *vertically*.
- At a summary-function node the score functions have to be added *horizontally*.
- At a scaling-function node the score function has to be scaled *horizontally and vertically*.

Summarizing, the message function (or equivalently the score function or the current-voltage characteristic) represents the influence of one part of an FFG (network) on the other part of the FFG (network), see Fig. 4.8.

Note that in the case of jointly Gaussian random variables the score functions are linear functions where

- the value where the voltage crosses the (horizontal) voltage-axis corresponds to the mean,
- the slope corresponds to the inverse variance, i.e., the steeper the slope the higher the certainty of the message is.

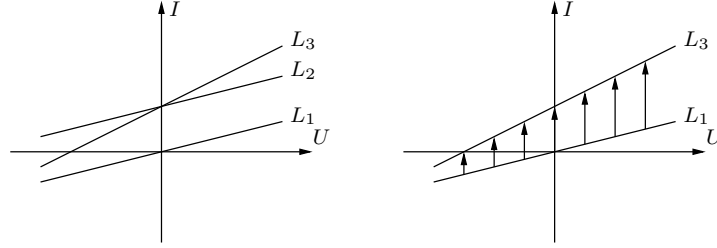


Figure 4.9: Left: The straight line  $L_3$  is the vertical addition of the straight lines  $L_1$  and  $L_2$ . Right: interpreting the vertical addition of the line  $L_2$  to the the line  $L_1$  as an affine mapping from  $L_1$  to  $L_3$ .

### 4.3 Matrix-Method for Updating Messages

The method we present in this section works only for the case of jointly Gaussian random variables. We take advantage of the fact that the score function of all initial messages are linear functions and that the update rules modify linear functions in an affine way. Therefore all messages are linear functions.

#### 4.3.1 Example (Equality-Function Node)

Consider an equality-function node of an FFG as given in Fig. 4.5(left). From the messages  $\mu_1(x_1)$  and  $\mu_2(x_2)$  we would like to calculate  $\mu_3(x_3)$ . As we have seen in Sec. 4.2.2, the update rule is

$$s_3(x_3) = s_1(x_3) + s_2(x_3), \quad (4.35)$$

where  $s_1(\cdot)$ ,  $s_2(\cdot)$ , and  $s_3(\cdot)$  are the score functions of the messages  $\mu_1(\cdot)$ ,  $\mu_2(\cdot)$ , and  $\mu_3(\cdot)$ , respectively. In terms of current-voltage characteristics this is

$$I_3(U_3) = I_1(U_3) + I_2(U_3). \quad (4.36)$$

Geometrically speaking, we have to add the two curves  $I_1(\cdot)$  and  $I_2(\cdot)$  (or  $s_1(\cdot)$  and  $s_2(\cdot)$ ) vertically. As mentioned above, in the case of jointly Gaussian random variables, all current-voltage characteristics (all score functions) are linear functions. The graph of  $I_1(\cdot)$ , i.e. the point set  $L_1 := \{(U, I) : U \in \mathbb{R}, I = I_1(U)\}$ , is a straight line and can be described in a parametrized way as  $L_1 = \{(U_1(\tau), I_1(\tau)) = (k_{11}\tau + k_{12}, k_{21}\tau + k_{22}) : \tau \in \mathbb{R}\}$  for some constants  $k_{11}$ ,  $k_{12}$ ,  $k_{21}$ ,  $k_{22}$ . Assume that  $I_2(U_2)$  is given in the form  $I_2(U_2) = G_2U_2 + I_{q2}$ . The addition in Eq. (4.36) is then in geometrical terms a modification of the straight line  $L_1$  to the straight line

$$L_3 := \{(U, I) : U \in \mathbb{R}, I = I_3(U)\} = \{(U_3(\tau), I_3(\tau)) : \tau \in \mathbb{R}\}, \quad (4.37)$$

where

$$U_3(\tau) := U_1(\tau), \quad (4.38)$$

$$I_3(\tau) := I_1(\tau) + G_2U_1(\tau) + I_{q2}, \quad (4.39)$$

or, rewritten,

$$\begin{pmatrix} U_3(\tau) \\ I_3(\tau) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ G_2 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \end{pmatrix} + \begin{pmatrix} 0 \\ I_{q2} \end{pmatrix}. \quad (4.40)$$

For an illustration, see Fig. 4.9. The matrix multiplication describes a scaling operation, whereas the vector addition is a translation.<sup>7</sup> But instead of having a matrix multiplication *and* a vector addition we would prefer to have only a matrix multiplication in order to simplify the concatenation of several such operations. To achieve this goal we use a well-known trick: we embed this two-dimensional real space in a three dimensional real space. If the the former space does not go through the origin of the latter, the operations of scaling and translation can both be expressed as scalings. To be specific, we embed the  $(U, I)$ -plane in a three-dimensional space  $(x, y, z)$  so that the  $(U, I)$ -plane is parallel to the  $(x, y)$ -plane with  $z = 1$ .<sup>8</sup> The reformulated Eq. (4.40) reads then

$$\begin{pmatrix} U_3(\tau) \\ I_3(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ G_2 & 1 & I_q \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix}. \quad (4.41)$$

To conclude this subsection we remark that there are also other possibilities to describe a straight line; e.g. every straight line in the  $(U, I)$ -plane is the solution of  $AU + BI + C = 0$  for some constants  $A$ ,  $B$ , and  $C$ . One would then have to consider how to update these three constants in the different function nodes.

### 4.3.2 The General Case

Applying the principle we introduced in Sec. 4.3.1 to the different function nodes, we obtain the following list.

- **(Equality-Function Node)** Assume that  $I_2(U_2) = G_2U_2 + I_q$ . We reformulate Eq. (4.13) to

$$\begin{pmatrix} U_3(\tau) \\ I_3(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ G_2 & 1 & I_q \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix}. \quad (4.42)$$

- **(Summation-Function Node)** Assume that  $U_2(I_2) = R_2I_2 + U_q$ . We reformulate Eq. (4.28) to

$$\begin{pmatrix} U_3(\tau) \\ I_3(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & R_2 & U_q \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix}. \quad (4.43)$$

<sup>7</sup>Note that  $U_1(\tau)$  and  $I_1(\tau)$  actually need not to be linear, only  $L_2$  has to be a straight line. But we are not pursuing this point further.

<sup>8</sup>Instead of this embedding of a two-dimensional space in a three-dimensional space we could have embedded the (affine) two-dimensional space in the projective two-dimensional space: the resulting equations would have been the same.



- **(Scaling-Function Node)** We reformulate Eq. (4.34) to

$$\begin{pmatrix} U_2(\tau) \\ I_2(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix}. \quad (4.44)$$

The next subsection shows how these rules can be used repeatedly.

### 4.3.3 The Matrix-Method for Updating Messages of the Scalar Kalman Filter

To exemplify these calculations, we consider the messages on the main horizontal processing edges of the Kalman filter FFG, see Fig. 2.9 with details in Fig. 2.10. Assume the message at edge  $X[k-1]$  is given by the straight line  $L_1 := \{(U_1(\tau), I_1(\tau)) : \tau \in \mathbb{R}\}$ . Then<sup>9</sup>

$$\begin{pmatrix} U_2(\tau) \\ I_2(\tau) \\ 1 \end{pmatrix} = \mathbf{T}_{2 \leftarrow 1} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} a[k] & 0 & 0 \\ 0 & a[k]^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix}. \quad (4.45)$$

The message coming from the  $b[k]$ -scaling function node to the summation-function node is  $U(I) = b[k]^2 \sigma_{U[k]}^2 I + b[k] m_{U[k]}$ :

$$\begin{pmatrix} U_3(\tau) \\ I_3(\tau) \\ 1 \end{pmatrix} = \mathbf{T}_{3 \leftarrow 2} \cdot \begin{pmatrix} U_2(\tau) \\ I_2(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & b[k]^2 \sigma_{U[k]}^2 & b[k] m_{U[k]} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_2(\tau) \\ I_2(\tau) \\ 1 \end{pmatrix}. \quad (4.46)$$

The message coming from the  $c[k]$ -scaling function node to the summation-function node is  $I(U) = c[k]^2 \sigma_{W[k]}^{-2} U + c[k] \sigma_{W[k]}^{-2} (m_{W[k]} - y[k])$ :

$$\begin{pmatrix} U_4(\tau) \\ I_4(\tau) \\ 1 \end{pmatrix} = \mathbf{T}_{4 \leftarrow 3} \cdot \begin{pmatrix} U_3(\tau) \\ I_3(\tau) \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ c[k]^2 \sigma_{W[k]}^{-2} & 1 & c[k] \sigma_{W[k]}^{-2} (m_{W[k]} - y[k]) \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} U_3(\tau) \\ I_3(\tau) \\ 1 \end{pmatrix}. \quad (4.47)$$

The whole iteration through a Kalman filter section thus can be summarized as

$$\begin{pmatrix} U_4(\tau) \\ I_4(\tau) \\ 1 \end{pmatrix} = \mathbf{T}_{4 \leftarrow 3} \cdot \mathbf{T}_{3 \leftarrow 2} \cdot \mathbf{T}_{2 \leftarrow 1} \cdot \begin{pmatrix} U_1(\tau) \\ I_1(\tau) \\ 1 \end{pmatrix}. \quad (4.48)$$

### 4.3.4 The Higher-Dimensional Case

Until now we assumed the edge variable  $X$  to be a scalar, so  $U$  and  $I$  were scalars. But the method can be carried over to the vector case.<sup>10</sup> Assume that the characteristic  $\mathbf{I}_1(\mathbf{U}_1)$  is described by the plane  $L_1 := \{(\mathbf{U}_1(\boldsymbol{\tau}), \mathbf{I}_1(\boldsymbol{\tau})) : \boldsymbol{\tau} \in \mathbb{R}^n\}$ . The rules of Sec. 4.3.2 get modified to<sup>11</sup>

<sup>9</sup>Note that  $U$  or  $U(\tau)$  are voltages whereas  $U[k]$  is the random variable denoting the SKF input.

<sup>10</sup>We use  $\mathbf{I}$  to denote the current vector and  $\mathbb{1}$  to denote the identity matrix. Additionally, note that  $\mathbf{U}$  or  $\mathbf{U}(\boldsymbol{\tau})$  are voltage vectors (voltage vector function), whereas  $\mathbf{U}[k]$  is the random vector of denoting the Kalman filter input and the edge in the FFG.

<sup>11</sup>If the matrix is not square or not invertible, one takes the Moore-Penrose generalized inverse as defined in App. F.

- **(Equality-Function Node)** Assume that  $\mathbf{I}_2(\mathbf{U}_2) = \mathbf{G}_2\mathbf{U}_2 + \mathbf{i}_q$ . Then

$$\begin{pmatrix} \mathbf{U}_3(\boldsymbol{\tau}) \\ \mathbf{I}_3(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{G}_2 & \mathbb{1} & \mathbf{i}_q \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\tau}) \\ \mathbf{I}_1(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.49)$$

- **(Summation-Function Node)** Assume that  $\mathbf{U}_2(\mathbf{I}_2) = \mathbf{R}_2\mathbf{I}_2 + \mathbf{U}_q$ . Then

$$\begin{pmatrix} \mathbf{U}_3(\boldsymbol{\tau}) \\ \mathbf{I}_3(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbf{R}_2 & \mathbf{U}_q \\ \mathbf{0} & \mathbb{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\tau}) \\ \mathbf{I}_1(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.50)$$

- **(Scaling-Function Node)** We consider scaling by  $\mathbf{A}$ . Then

$$\begin{pmatrix} \mathbf{U}_2(\boldsymbol{\tau}) \\ \mathbf{I}_2(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\tau}) \\ \mathbf{I}_1(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.51)$$

#### 4.3.5 The Matrix-Method for Updating Messages of the Vector Kalman Filter

The update rules for the scalar Kalman filter in Sec. 4.3.3 can be reformulated to get the vector Kalman filter. Assume the message at edge  $\mathbf{X}[k-1]$  is given by the plane  $L_1 := \{(\mathbf{U}_1(\boldsymbol{\tau}), \mathbf{I}_1(\boldsymbol{\tau})) : \boldsymbol{\tau} \in \mathbb{R}^n\}$ , where  $n$  is the dimension of  $\mathbf{X}[k-1]$ . Then

$$\begin{pmatrix} \mathbf{U}_2(\boldsymbol{\tau}) \\ \mathbf{I}_2(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \mathbf{T}_{2 \leftarrow 1} \cdot \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\tau}) \\ \mathbf{I}_1(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{A}[k] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}[k]^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\tau}) \\ \mathbf{I}_1(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.52)$$

The message coming from the  $\mathbf{B}[k]$ -scaling function node to the summation-function node is  $\mathbf{U}(\mathbf{I}) = \mathbf{B}[k]\mathbf{K}_{\mathbf{U}[k]}\mathbf{B}[k]^T\mathbf{I} + \mathbf{B}[k]\mathbf{m}_{\mathbf{U}[k]}$ :

$$\begin{pmatrix} \mathbf{U}_3(\boldsymbol{\tau}) \\ \mathbf{I}_3(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \mathbf{T}_{3 \leftarrow 2} \cdot \begin{pmatrix} \mathbf{U}_2(\boldsymbol{\tau}) \\ \mathbf{I}_2(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbf{B}[k]\mathbf{K}_{\mathbf{U}[k]}\mathbf{B}[k]^T & \mathbf{B}[k]\mathbf{m}_{\mathbf{U}[k]} \\ \mathbf{0} & \mathbb{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}_2(\boldsymbol{\tau}) \\ \mathbf{I}_2(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.53)$$

The message coming from the  $\mathbf{C}[k]$ -scaling function node to the summation-function node is  $\mathbf{I}(\mathbf{U}) = \mathbf{C}[k]^T\mathbf{K}_{\mathbf{W}[k]}^{-1}\mathbf{C}[k]\mathbf{U} + \mathbf{C}[k]^T\mathbf{K}_{\mathbf{W}[k]}^{-1}(\mathbf{m}_{\mathbf{W}[k]} - \mathbf{y}[k])$ :

$$\begin{pmatrix} \mathbf{U}_4(\boldsymbol{\tau}) \\ \mathbf{I}_4(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \mathbf{T}_{4 \leftarrow 3} \cdot \begin{pmatrix} \mathbf{U}_3(\boldsymbol{\tau}) \\ \mathbf{I}_3(\boldsymbol{\tau}) \\ 1 \end{pmatrix} \quad (4.54)$$

$$= \begin{pmatrix} \mathbb{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{C}[k]^T\mathbf{K}_{\mathbf{W}[k]}^{-1}\mathbf{C}[k]\mathbf{U} & \mathbb{1} & \mathbf{C}[k]^T\mathbf{K}_{\mathbf{W}[k]}^{-1}(\mathbf{m}_{\mathbf{W}[k]} - \mathbf{y}[k]) \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{U}_3(\boldsymbol{\tau}) \\ \mathbf{I}_3(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.55)$$

The whole iteration through a Kalman filter section thus can be summarized as

$$\begin{pmatrix} \mathbf{U}_4(\boldsymbol{\tau}) \\ \mathbf{I}_4(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \mathbf{T}_{4 \leftarrow 3} \cdot \mathbf{T}_{3 \leftarrow 2} \cdot \mathbf{T}_{2 \leftarrow 1} \cdot \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\tau}) \\ \mathbf{I}_1(\boldsymbol{\tau}) \\ 1 \end{pmatrix}. \quad (4.56)$$

### 4.3.6 Renormalization

What remains to consider is how to start the whole process and how to extract at the end the desired information.

There are two practical ways to start the process. If the first function node is an equality-function node, the plane  $L_1$  should be the  $\mathbf{U}$ -axis.<sup>12</sup> Parametrized with  $\boldsymbol{\tau}$ , the vector case reads

$$\mathbf{U}(\boldsymbol{\tau}) := k \cdot \boldsymbol{\tau}, \quad (4.57)$$

$$\mathbf{I}(\boldsymbol{\tau}) := \mathbf{0}. \quad (4.58)$$

for an arbitrary scalar constant  $k$  (with correct dimension). If the first function node is a summation function node, the plane  $L_1$  should be the  $\mathbf{I}$ -axis.<sup>13</sup> Parametrized with  $\boldsymbol{\tau}$ , the vector case reads

$$\mathbf{U}(\boldsymbol{\tau}) := \mathbf{0}, \quad (4.59)$$

$$\mathbf{I}(\boldsymbol{\tau}) := k \cdot \boldsymbol{\tau}. \quad (4.60)$$

for an arbitrary scalar constant  $k$  (with correct dimension).

Suppose now that at some point we get

$$\begin{pmatrix} \mathbf{U}'(\boldsymbol{\tau}) \\ \mathbf{I}'(\boldsymbol{\tau}) \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 & \mathbf{m}_3 \\ \mathbf{M}_4 & \mathbf{M}_5 & \mathbf{m}_6 \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{U}(\boldsymbol{\tau}) \\ \mathbf{I}(\boldsymbol{\tau}) \\ 1 \end{pmatrix}, \quad (4.61)$$

and that the initialization was  $\mathbf{U}(\boldsymbol{\tau}) := \mathbf{0}, \mathbf{I}(\boldsymbol{\tau}) := k \cdot \boldsymbol{\tau}$  as in Eq. (4.58) so that

$$\begin{pmatrix} \mathbf{U}' \\ \mathbf{I}' \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 & \mathbf{m}_3 \\ \mathbf{M}_4 & \mathbf{M}_5 & \mathbf{m}_6 \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ k \cdot \boldsymbol{\tau} \\ 1 \end{pmatrix}, \quad (4.62)$$

which is equivalent to

$$\mathbf{U}' = k\mathbf{M}_2\boldsymbol{\tau} + \mathbf{m}_3, \quad (4.63)$$

$$\mathbf{I}' = k\mathbf{M}_5\boldsymbol{\tau} + \mathbf{m}_6. \quad (4.64)$$

We would like to eliminate the parameters  $\boldsymbol{\tau}$ . From Eq. (4.64) we get<sup>14</sup>

$$\mathbf{U}'(\mathbf{I}') = \mathbf{M}_2\mathbf{M}_5^{-1}\mathbf{I}' + \mathbf{m}_3 - \mathbf{M}_2\mathbf{M}_5^{-1}\mathbf{m}_6, \quad (4.65)$$

$$\mathbf{I}'(\mathbf{U}') = \mathbf{M}_5\mathbf{M}_2^{-1}\mathbf{U}' + \mathbf{m}_6 - \mathbf{M}_5\mathbf{M}_2^{-1}\mathbf{m}_3. \quad (4.66)$$

To find the estimate we simple have to evaluate  $\mathbf{U}'(\mathbf{0})$  (the “zero-crossing point” of the inverted score function) and the error covariance matrix is  $\mathbf{M}_2\mathbf{M}_5^{-1}$  (the “slope” of the inverted score function).<sup>15</sup>

<sup>12</sup>This plane guarantees that  $\mathbf{I}_3(\mathbf{U}_3) = \mathbf{I}_2(\mathbf{U}_3)$ .

<sup>13</sup>This plane guarantees that  $\mathbf{U}_3(\mathbf{I}_3) = \mathbf{U}_2(\mathbf{I}_3)$ .

<sup>14</sup>Note that by eliminating  $\boldsymbol{\tau}$ , the parameter  $k$  disappears also. This must be so, as it was anyway arbitrary.

<sup>15</sup>This procedure is also useful in bringing the plane description into the “standard forms” required in Sec. 4.3.4 for the description of  $\mathbf{I}_2(\mathbf{U}_2)$  and  $\mathbf{U}_2(\mathbf{I}_2)$ , respectively.

In general, to find the matrices and vectors in Eq. (4.65) we need to compute  $\mathbf{M}_2\mathbf{M}_5^{-1}$ ; whereas to evaluate the matrices and vectors in Eq. (4.66) we need to compute  $\mathbf{M}_5\mathbf{M}_2^{-1}$ . We now focus on the former case and show how the value of  $\mathbf{M}_2\mathbf{M}_5^{-1}$  can be propagated so that inverting large matrices can sometimes be avoided. This is e.g. the case in Kalman filtering when the measured output has a lower dimension than the state space. Let  $\mathbf{M} := \mathbf{M}_2\mathbf{M}_5^{-1}$  and see how this matrix has to be updated in the various function nodes.

- **(Equality-Function Node)**

$$\begin{pmatrix} \times & \mathbf{M}'_2 & \times \\ \times & \mathbf{M}'_5 & \times \\ \times & \times & \times \end{pmatrix} := \begin{pmatrix} \mathbb{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{G} & \mathbb{1} & \mathbf{i}_q \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \times & \mathbf{M}_2 & \times \\ \times & \mathbf{M}_5 & \times \\ \times & \mathbf{0} & \times \end{pmatrix} = \begin{pmatrix} \times & \mathbf{M}_2 & \times \\ \times & \mathbf{GM}_2 + \mathbf{M}_5 & \times \\ \times & \times & \times \end{pmatrix}. \quad (4.67)$$

We conclude that

$$\mathbf{M}' := \mathbf{M}'_2\mathbf{M}'_5^{-1} = \mathbf{M}_2(\mathbf{GM}_2 + \mathbf{M}_5)^{-1} = \mathbf{M}_2\mathbf{M}_5^{-1}(\mathbf{GM}_2\mathbf{M}_5^{-1} + \mathbb{1})^{-1} \quad (4.68)$$

$$= \mathbf{M}(\mathbb{1} + \mathbf{GM})^{-1}. \quad (4.69)$$

Assume that  $\mathbf{G} := \mathbf{C}^T\mathbf{K}^{-1}\mathbf{C}$  as in Kalman filtering. Using the Matrix-Inversion Lemma (see e.g. p. 656 of [15])

$$(\mathbf{H}_1 + \mathbf{H}_2\mathbf{H}_3\mathbf{H}_4)^{-1} = \mathbf{H}_1^{-1} - \mathbf{H}_1^{-1}\mathbf{H}_2(\mathbf{H}_4\mathbf{H}_1^{-1}\mathbf{H}_2 + \mathbf{H}_3^{-1})\mathbf{H}_4\mathbf{H}_1^{-1}, \quad (4.70)$$

where  $\mathbf{H}_1$  and  $\mathbf{H}_2$  must be non-singular square matrices (not necessarily of the same size), we get (with  $\mathbf{H}_1 := \mathbb{1}$ ,  $\mathbf{H}_2 := \mathbf{C}^T$ ,  $\mathbf{H}_3 := \mathbf{K}^{-1}$ , and  $\mathbf{H}_4 := \mathbf{CM}$ )

$$\mathbf{M}' = \mathbf{M}(\mathbb{1} + \mathbf{C}^T\mathbf{K}^{-1}\mathbf{CM})^{-1} = \mathbf{M}[\mathbb{1} - \mathbb{1}\mathbf{C}^T(\mathbf{CM}\mathbb{1}\mathbf{C}^T + \mathbf{K})^{-1}\mathbf{CM}\mathbb{1}] \quad (4.71)$$

$$= \mathbf{M} - \mathbf{MC}^T(\mathbf{CMC}^T + \mathbf{K})^{-1}\mathbf{CM}, \quad (4.72)$$

a formula which is especially easy in the case of scalar measurements as then the inversion turns out to be a scalar inversion.

- **(Summation-Function Node)**

$$\begin{pmatrix} \times & \mathbf{M}'_2 & \times \\ \times & \mathbf{M}'_5 & \times \\ \times & \times & \times \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbf{R} & \mathbf{U}_q \\ \mathbf{0} & \mathbb{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \times & \mathbf{M}_2 & \times \\ \times & \mathbf{M}_5 & \times \\ \times & \mathbf{0} & \times \end{pmatrix} = \begin{pmatrix} \times & \mathbf{M}_2 + \mathbf{RM}_5 & \times \\ \times & \mathbf{M}_5 & \times \\ \times & \times & \times \end{pmatrix}. \quad (4.73)$$

We conclude that

$$\mathbf{M}' := \mathbf{M}'_2\mathbf{M}'_5^{-1} = (\mathbf{M}_2 + \mathbf{RM}_5)\mathbf{M}_5^{-1} = \mathbf{M}_2\mathbf{M}_5^{-1} + \mathbf{R} = \mathbf{M} + \mathbf{R}. \quad (4.74)$$

- **(Scaling-Function Node)** We consider scaling by  $\mathbf{A}$ . Then

$$\begin{pmatrix} \times & \mathbf{M}'_2 & \times \\ \times & \mathbf{M}'_5 & \times \\ \times & \times & \times \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \times & \mathbf{M}_2 & \times \\ \times & \mathbf{M}_5 & \times \\ \times & \mathbf{0} & \times \end{pmatrix} = \begin{pmatrix} \times & \mathbf{AM}_2 & \times \\ \times & \mathbf{A}^{-T}\mathbf{M}_5 & \times \\ \times & \times & \times \end{pmatrix}. \quad (4.75)$$

We conclude that

$$\mathbf{M}' := \mathbf{M}'_2\mathbf{M}'_5^{-1} = \mathbf{AM}_2(\mathbf{A}^{-T}\mathbf{M}_5)^{-1} = \mathbf{AM}_2\mathbf{M}_5^{-1}\mathbf{A}^T = \mathbf{AMA}^T. \quad (4.76)$$

We remark that the update of  $\mathbf{M}$  to  $\mathbf{M}'$  is only complicated in the first case whereas it is very straightforward in the second and third cases. The update rules for  $\mathbf{M}_5\mathbf{M}_2^{-1}$  would be derived similarly, but now the summation-function node gives the ‘‘complicated rule’’.

## 4.4 Some Conclusions About the Score Function

It seems that the score function is very suitable for updating the messages of the MPA (for any random variables!). The proof with the help of electrical networks was especially simple. And the resulting rules are very simple geometrically and can be used to get some intuition about what is going on e.g. in a Kalman filter (see also the comments about the score function in the scalar case after Eq. (4.4)).

It is important that we have always assumed that the messages have a convex negative exponent, meaning that the score function is monotonically increasing. What happens in more general contexts, is one of the questions raised in Sec. 7.



## Chapter 5

# Miscellaneous Topics

This chapter discusses various topics related to the derivation of electrical networks from FFGs as shown in Ch. 3.

### 5.1 Nonlinear Resistors and a Dictionary between Probability Density Functions and $I(U)$ -characteristics

We have seen that a function node  $f(\cdot)$  as in Fig. 3.3 corresponds to a generalized resistor with characteristic

$$I = I(U) = s(U). \quad (5.1)$$

with the score function<sup>1</sup>

$$s(x) = \frac{\partial}{\partial x} (-\ln f(x)) = -\frac{\frac{\partial}{\partial x} f(x)}{f(x)}. \quad (5.2)$$

The  $I(U)$ -characteristic equals the score function of the pdf. On the other hand, given an  $I(U)$ -characteristic, the corresponding function is

$$f(x) \propto \exp\left(-\int_0^U I(U') dU'\right)\Big|_{U=x}. \quad (5.3)$$

If  $f(\cdot)$  is a density then the proportionality constant (which comes from the integration constant) can be chosen so as to guarantee  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

In general,  $s(\cdot)$  is non-linear. Then  $-\ln f(\cdot)$  is a (strictly) convex function if and only if  $s(\cdot)$  is a (strictly) increasing function. If  $f(\cdot)$  is a Gaussian distribution function, then  $s(\cdot)$  is a linear function and  $-\ln f(\cdot)$  is a quadratic function.

Tab. G.3 lists different distribution functions and their corresponding current-voltage characteristics.

### 5.2 The Vector Case

The vector case basically does not pose any additional problems compared to the scalar case. This is because the blockwise MAP estimate of the joint density maximizes over all

---

<sup>1</sup>See also Footnote 1 on page 29.

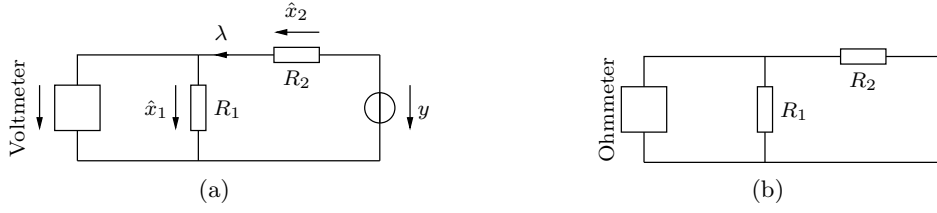


Figure 5.1: Measurement of the estimate  $\hat{x}_1$  in (a) and measurement of the error variance  $E[(X_1 - \hat{X}_1)^2]$  in (b) in the electrical network of Example 1 of Sec. 2.1 (see Fig. 2.2).

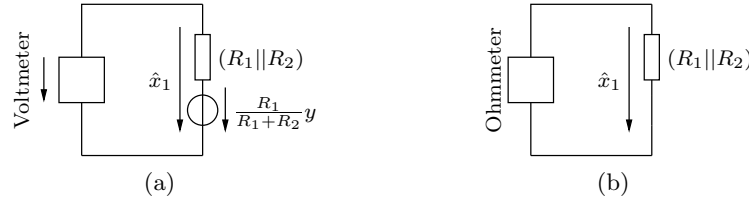


Figure 5.2: Measurement of the estimate  $\hat{x}_1$  in (a) and measurement of the expected squared error  $E[(X_1 - \hat{X}_1)^2]$  in (b) of the modified electrical network of Example 1 of Sec. 2.1 (see Fig. 2.4). Note that in the electrical network in part (a) no current flows.

arguments jointly. So, e.g. maximizing with respect to  $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$  is equivalent to maximizing with respect to  $((x_1, x_2), (x_3), (x_4, x_5, x_6), \dots, (x_n))$ .

In the case of jointly Gaussian random variables, the scalar MPA is equivalent to the scalar MSA and the vector MPA is equivalent to the vector MSA.

### 5.3 The Error Covariance Matrix in the Case of Jointly Gaussian Random Variables

In the case of jointly Gaussian random variables, not only the blockwise MAP estimates can be measured in the network, but also the the entries of the error covariance matrix can be measured. Because in the case of jointly Gaussian random variables the estimates are bias-free, i.e.  $E[X - \hat{X}] = 0$  (see e.g. App. A), we have

$$E[\{(X_1 - \hat{X}_1) - E[X_1 - \hat{X}_1]\}\{(X_2 - \hat{X}_2) - E[X_2 - \hat{X}_2]\}] = E[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2)]. \quad (5.4)$$

#### 5.3.1 Introductory Example

To start, we consider again the example from Sec. 2.1 where we had the estimation vector

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} \frac{R_1}{R_1 + R_2} y \\ \frac{R_2}{R_1 + R_2} y \end{pmatrix} \quad (5.5)$$

and error covariance matrix

$$\begin{pmatrix} E[(X_1 - \hat{X}_1)^2] & E[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2)] \\ E[(X_2 - \hat{X}_2)(X_1 - \hat{X}_1)] & E[(X_2 - \hat{X}_2)^2] \end{pmatrix} = \begin{pmatrix} +\frac{R_1 R_1}{R_1 + R_2} & -\frac{R_1 R_1}{R_1 + R_2} \\ -\frac{R_1 R_1}{R_1 + R_2} & +\frac{R_1 R_1}{R_1 + R_2} \end{pmatrix}. \quad (5.6)$$



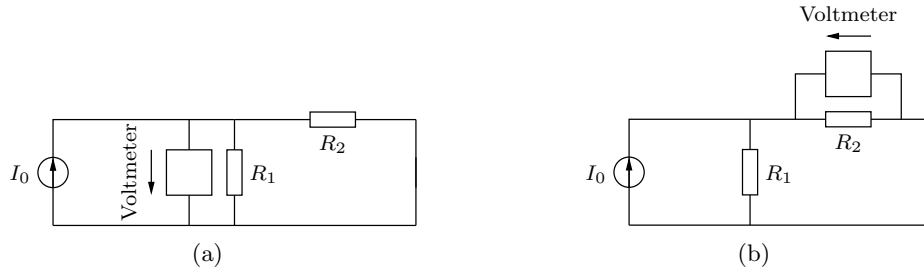


Figure 5.3: (a) Alternative way (compared to Fig. 2.2(b)) of doing the measurement needed for getting the expected squared error  $E[(X_1 - \hat{X}_1)^2]$ . (b) Measurement of the error covariance  $E[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2)]$ .

Fig. 5.1(a) shows what measurements have to be performed to get the estimate  $\hat{x}_1$ , whereas Fig. 5.1(b) shows the measurements needed to get  $E[(X_1 - \hat{X}_1)^2]$ . The first result follows from the way we set up the electrical network whereas the second result has to be proven additionally; this is the aim of this section. Note that for the second measurements all voltage sources have to be short-circuited and all current sources have to be opened. As long as we do not remove the nodes between which we have  $\hat{x}_1$ , we can modify<sup>2</sup> the network as we did at the end of Sec. 2.1. Performing then the measurements is nearly trivial as can be seen in Fig. 5.2. Finally, Fig. 5.3(a) shows how to apply a current source and a voltagemeter to measure  $E[(X_1 - \hat{X}_1)^2]$  and  $E[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2)]$ : the ratio of measured voltage divided by the applied current gives the desired value.

### 5.3.2 The General Case

After this introductory example we are ready to look at the general case. Assume that  $\mathbf{X}$  and  $\mathbf{Y}$  have the joint Gaussian density and can be written as<sup>3</sup>

$$p_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) \propto \exp \left( -\frac{1}{2} \begin{pmatrix} \mathbf{x} - \mathbf{m}_X \\ \mathbf{y} - \mathbf{m}_Y \end{pmatrix}^T \begin{pmatrix} \mathbf{K}_{XX} & \mathbf{K}_{XY} \\ \mathbf{K}_{XY}^T & \mathbf{K}_{YY} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x} - \mathbf{m}_X \\ \mathbf{y} - \mathbf{m}_Y \end{pmatrix} \right). \quad (5.7)$$

We define

$$\begin{pmatrix} \mathbf{D}_{XX} & \mathbf{D}_{XY} \\ \mathbf{D}_{XY}^T & \mathbf{D}_{YY} \end{pmatrix} := \begin{pmatrix} \mathbf{K}_{XX} & \mathbf{K}_{XY} \\ \mathbf{K}_{XY}^T & \mathbf{K}_{YY} \end{pmatrix}^{-1}, \quad (5.8)$$

and, as shown on p. 656 of [15], one has

$$\mathbf{D}_{XX} = (\mathbf{K}_{XX} - \mathbf{K}_{XY} \mathbf{K}_{YY}^{-1} \mathbf{K}_{XY}^T)^{-1}, \quad (5.9)$$

$$\mathbf{D}_{XY} = -\mathbf{D}_{XX} \mathbf{K}_{XY} \mathbf{K}_{YY}^{-1}. \quad (5.10)$$

<sup>2</sup>The statement about what modifications are allowed in the case of the measurement of diagonal entries of the error covariance matrix can be extended to what modifications are allowed in the case of the measurement of non-diagonal entries.

<sup>3</sup>Applying the results of Sec. 1.6 one can transform a distribution containing a jointly Gaussian density and some Dirac-delta terms into a jointly Gaussian density. The two densities are equal in the limit  $\beta \rightarrow 0$ .

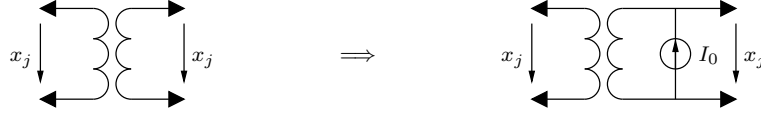


Figure 5.4: Changes that have to be applied to the partial electrical network corresponding to the variable edge  $X_j$ , where  $I_0 := 1$ . (Although it looks at first sight asymmetric, it is indeed symmetric.)

The blockwise MAP estimate  $\hat{\mathbf{x}}$  based on the measurement  $\mathbf{Y} = \mathbf{y}$  is

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg \max_{\mathbf{x}} p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \arg \max_{\mathbf{x}} p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}). \quad (5.11)$$

Assuming the joint density  $p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$  of Eq. (5.7), the conditional density  $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$  is again jointly Gaussian<sup>4</sup>

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}))^T \mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}^{-1} (\mathbf{x} - \mathbf{m}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}))\right) \quad (5.12)$$

with

$$\mathbf{m}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}) = \mathbf{m}_{\mathbf{X}} + \mathbf{K}_{\mathbf{X}\mathbf{Y}} \mathbf{K}_{\mathbf{Y}\mathbf{Y}}^{-1} (\mathbf{y} - \mathbf{m}_{\mathbf{Y}}), \quad (5.13)$$

$$\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}} = \mathbf{D}_{\mathbf{X}\mathbf{X}}^{-1} = \mathbf{K}_{\mathbf{X}\mathbf{X}} - \mathbf{K}_{\mathbf{X}\mathbf{Y}} \mathbf{K}_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{K}_{\mathbf{Y}\mathbf{X}}^T. \quad (5.14)$$

From this and the second arg max of Eq. (5.11) it is easily seen that the MAP estimate  $\hat{\mathbf{x}}$  is

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{m}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}), \quad (5.15)$$

and the error covariance matrix is<sup>5</sup>

$$\mathbb{E}[(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^T] = \mathbb{E}\left[\mathbb{E}\left[(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^T \middle| \mathbf{Y}\right]\right] = \mathbb{E}[\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}] = \mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}. \quad (5.16)$$

Now we would like to relate these results to the measurements as shown at the beginning of this section. Because the correctness of the measurement of the estimate is clear from the problem statement, we can focus on the measurement of the error covariance matrix. To analyze our networks we use the first arg max expression of Eq. (5.11). Maximizing  $p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$  is equivalent to minimizing  $-\ln(p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}))$  and setting its gradient with respect to  $\mathbf{x}$  equal to  $\mathbf{0}$

$$-\frac{\partial}{\partial \mathbf{x}} \ln(p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})) = \mathbf{D}_{\mathbf{X}\mathbf{X}}(\mathbf{x} - \mathbf{m}_{\mathbf{X}}) + \mathbf{D}_{\mathbf{X}\mathbf{Y}}(\mathbf{y} - \mathbf{m}_{\mathbf{Y}}) \stackrel{!}{=} \mathbf{0}. \quad (5.17)$$

This vector equation can be represented by an electrical network where every line corresponds to a Kirchhoff current law with voltage sources  $\mathbf{m}_{\mathbf{X}}$ ,  $\mathbf{y}$ , and  $\mathbf{m}_{\mathbf{Y}}$ . We introduce now the following twist:

- we set the gradient not equal to zero but equal to  $\mathbf{e}^{(j)}$ , a vector that has a one at position  $j$  and is zero otherwise, and

<sup>4</sup>To emphasize that  $\mathbf{m}_{\mathbf{X}|\mathbf{Y}}(\mathbf{y})$  is a function of  $\mathbf{y}$ , whereas  $\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}$  is not, we added  $\mathbf{y}$  as an argument in the former case.

<sup>5</sup>See also Eq. (5.4) and the comments preceding it.

- set all sources (i.e.  $\mathbf{m}_X$ ,  $\mathbf{y}$ , and  $\mathbf{m}_Y$ ) equal to zero.

We get

$$\mathbf{D}_{\mathbf{X}\mathbf{X}} \cdot \mathbf{x} \stackrel{!}{=} \mathbf{e}^{(j)}, \quad (5.18)$$

or equivalently,

$$\mathbf{x} = \mathbf{D}_{\mathbf{X}\mathbf{X}}^{-1} \cdot \mathbf{e}^{(j)} = \mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}} \cdot \mathbf{e}^{(j)}. \quad (5.19)$$

The voltage  $x_i$  is therefore the value  $[\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}]_{i,j}$  the entry in row  $i$  and column  $j$  of  $\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}$ . The next step is to find out what this twist means for the electrical networks derived in Ch. 3. We assume that the edge  $X_j$  in the FFG is connected to two constraint function nodes (see Sec. 3.2.1). The other cases follow in the same style. We must reformulate the whole problem by writing a joint density of all involved variables instead of a density plus constraints. This can be done as pointed out in Sec. 1.6.  $\delta(x - x')$ -terms are replaced by  $\text{const} \cdot \exp(-\frac{1}{2\beta}(x - x')^2)$ -terms and the Lagrangian  $L = -\ln p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$  contains terms of the form  $+\frac{1}{2\beta}(x - x')^2$ . The development in Ch. 3 can be parallelized in this new formulation. In the limit  $\beta \rightarrow 0$  one obtains the same partial electrical networks. The above twist in this context now means the following.

- Setting the gradient of the Lagrangian  $L$  not equal to zero but equal to  $\mathbf{e}^{(j)}$  does only change the partial electrical related to the edge  $X_j$ : this part has to be changed as shown in Fig. 5.4.
- Setting  $\mathbf{m}_X$ ,  $\mathbf{y}$ , and  $\mathbf{m}_Y$  equal to zero, means to set all other sources equal to zero.

Measuring the voltage corresponding to edge  $X_i$  then gives the desired entry  $[\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}]_{i,j}$  of  $\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}$ . We conclude with three remarks.

- For any error covariance matrix we have  $[\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}]_{i,j} = [\mathbf{K}_{\mathbf{X}\mathbf{X}|\mathbf{Y}}]_{j,i}$  for any  $i$  and  $j$ . But this is nothing else than an expression of Green's reciprocity theorem for networks of linear resistors (see e.g. Ch. 13.3 of [16]): applying a current at a first place and measuring the voltage at a second place gives the same voltage-current ratio as applying the current at the second place and measuring the voltage at the first place.
- Assume that the electrical network contains only ideal voltage sources and no ideal current sources. For the estimates, the absolute values of the resistors do not matter as long as the ratios between the different resistors remains the same. This is so because the estimates are voltages and are only functions of the ideal voltage sources and the ratios of the resistors. But for the measurement of the entries of the covariance matrix one considers voltages divided by currents; so here the absolute values of the resistors matter. Therefore, the electrical networks as derived in Ch. 3 are in this sense canonical and to be preferred to other "equivalent" networks where the resistors have been rescaled.
- Note that for the above measurements we did not have to change the electrical network. Earlier we stated that for measuring entries of the error covariance matrix we have to set the sources to zero. This is not necessary: we could have let the voltage sources on and the desired ratio would have been the measured voltage change divided by the applied current change.

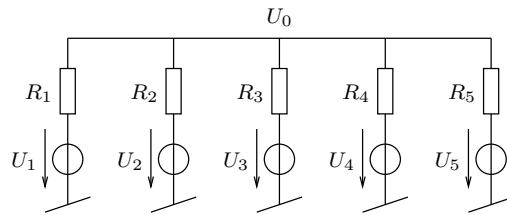


Figure 5.5: Electrical network for weighted averaging: the potential  $U_0$  is the weighted average of  $U_1$  to  $U_5$  weighted according to  $1/R_i$ ,  $i = 1, \dots, 5$ .

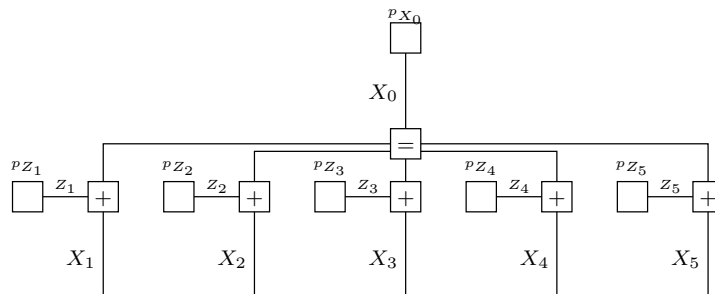


Figure 5.6: FFG to the electrical network in Fig. 5.5.

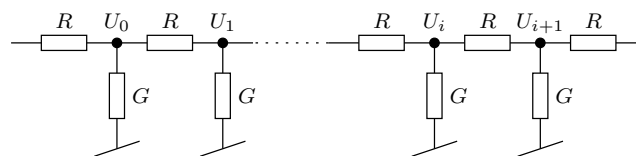


Figure 5.7: Chain of resistors.

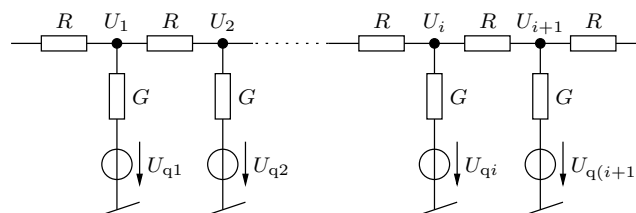


Figure 5.8: Chain of resistors with voltage sources.

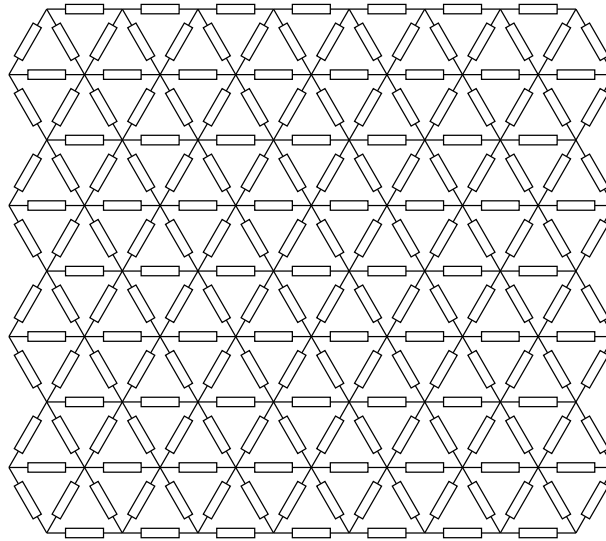


Figure 5.9: Topology of a hexagonal network. Each node is connected through a resistor and a voltage source in series to ground (or possibly through a more complicated partial electrical network).

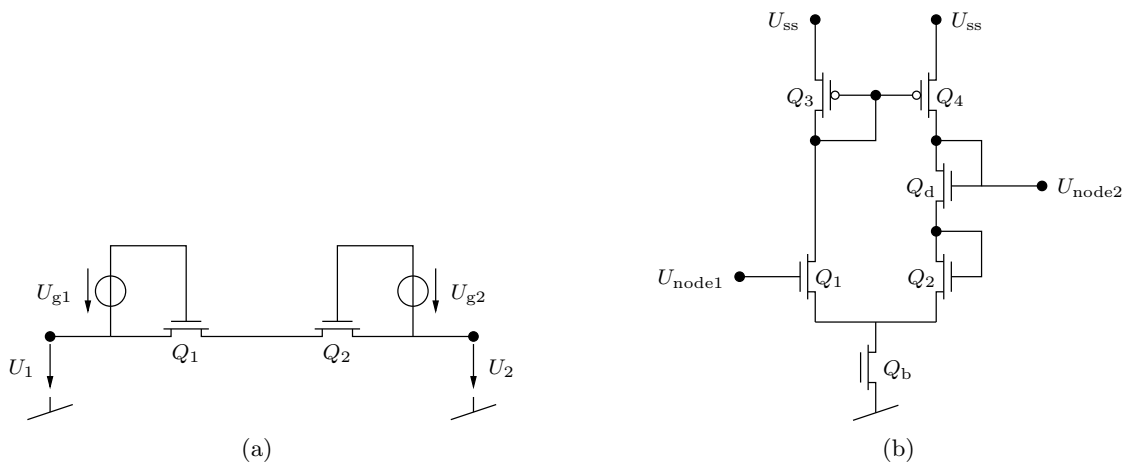


Figure 5.10: (a) shows a tanh-resistor. (b) shows a network needed twice in part (a) for generating  $U_g$ . ( $U_g$  is available between  $U_{node2}$  and  $U_{node1}$ .)

## 5.4 Examples from the Book by Mead

As mentioned in the introduction to this report, one source of inspiration were the electrical networks in the book by Mead [4] where several simple but efficient electrical networks are presented that perform specific tasks. The common idea behind all these circuits is to use the characteristics of available simple electrical elements (especially in VLSI). Putting together lots of simple elements results in circuits inspired by nature that operate highly parallelly. Using the approach taken in this report, these can now be reinterpreted in a probabilistic and optimizing sense. The aim of this section is to have a quick look at some of the circuits in [4].

- We start with Fig. 5.5 (see also Fig. 7.2 of [4]), which represents a network for averaging  $U_1$  to  $U_5$  with weights  $1/R_i$ , i.e. the potential  $U_0$  is

$$U_0 = \frac{\sum_{i=1}^5 \frac{1}{R_i} U_i}{\sum_{i=1}^5 \frac{1}{R_i}}. \quad (5.20)$$

The probabilistic interpretation is given by the FFG in Fig. 5.6: given the measurements  $X_1 = x_1, \dots, X_5 = x_5$ , we want to find the (blockwise) MAP estimate of  $X_0$ . Starting from this FFG and using the methods of Ch. 3 we would have obtained the electrical network in Fig. 5.5 with the following densities:  $Z_1 \sim \mathcal{N}(0, R_1), \dots, Z_5 \sim \mathcal{N}(0, R_5)$ , and  $X_0 \sim \mathcal{N}(0, \sigma_0^2)$  where  $\sigma_0^2 \rightarrow \infty$ . Changing the different distributions (see e.g. Sec. 5.1 and Tab. G.3), one can obtain other electrical networks.

- We turn now to Figs. 5.7 and 5.8 (see also Figs. 7.3, 7.5, and C.5 of [4]): The first models some propagation of a wave when a voltage is applied e.g. on the left. The second models a type of local averaging of the values of the voltage sources  $U_{q,i}$ , which appears at the nodes  $U_i$ . This can now be interpreted as Kalman filters, or, as if they are space-invariant, as Wiener filters.
- Related to these circuits is the one shown in Fig. 5.9 (see also Figs. 7.6 and 15.2 of [4]), where additionally each node is connected through a resistor and a voltage source in series to ground (or possibly through a more complicated electrical network). As above, this circuit does a type of local averaging of the value of the voltage sources. This circuit models a two-dimensional filtering as it happens in the retina of an eye. Again, this circuit can be interpreted as a Kalman filter, or as Wiener filter (if it the resistors are space-invariant).
- Using resistors whose current-voltage characteristic has a tanh-shape can be advantageous for segmentation tasks as shown in Fig. 7.14 of [4]. As such non-linear resistors are quite easily implemented (see Fig. 7.9 and 7.10 of [4], or Fig. 5.10 of this report), they can be used for applications where lots of such elements are needed. In the interpretation of probabilities, the use of such generalized resistors corresponds to non-Gaussian random variables.
- Electrical networks modeling a delay line as in Fig. 5.12 (see also Fig. 9.4 of [4]) are interpreted in Sec. 5.6.

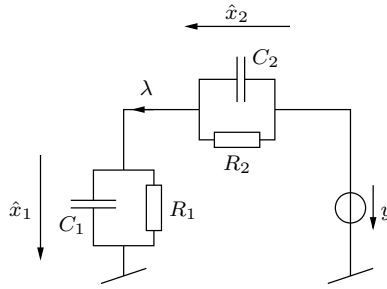


Figure 5.11: Gradient search electrical network.

## 5.5 Capacitors and Inductors

Up to now all circuits were time-independent as all component characteristics were static. When capacitors are introduced, we observe that time-dependence can occur in two different contexts.

### 5.5.1 Complex Random Variables

A complex jointly Gaussian random vector  $\mathbf{X}$  has the density<sup>6</sup>

$$p_{\mathbf{X}}(\mathbf{x}) \propto \exp\left(-(\mathbf{x} - \mathbf{m}_{\mathbf{X}})^H \mathbf{K}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x} - \mathbf{m}_{\mathbf{X}})\right), \quad (5.21)$$

where  $\mathbf{K}_{\mathbf{X}\mathbf{X}}$  is Hermitian, i.e.,  $\mathbf{K}_{\mathbf{X}\mathbf{X}}^H = \mathbf{K}_{\mathbf{X}\mathbf{X}}$ . Such random variables have as properties e.g. that  $E[\Re(X_i)^2] = E[\Im(X_i)^2] = [\mathbf{K}_{\mathbf{X}\mathbf{X}}]_{i,i}/2$  and  $E[\Re(X_i)\Im(X_i)] = 0$ .

Of course, every complex random variable can be seen as a vector of two real random variables. The possible densities for complex random variables (vectors) are therefore a subset of the possible densities for real random variables (vectors).

The theory of Kalman filtering is not changed by considering complex jointly Gaussian random variables, the only change necessary is to replace the transposition operator for matrices by the Hermitian operator for matrices and the inner product of  $X_1$  and  $X_2$  is not  $E[X_1 X_2]$  but  $E[X_1 \overline{X_2}]$  (see e.g. [8]). So, basically nothing new happens; but as complex analysis is “more beautiful” than real analysis, special things can also happen here.

The derivations done in Ch. 3 can be carried over to complex random vectors. In the case of complex jointly Gaussian random vectors we obtain a system of linear equation in complex variables. This system of equations can be implemented with a circuit containing resistors, capacitors, inductors, ideal sinusoidal voltage sources, ideal sinusoidal current sources, and impedance converters (or transformers). Fixing a frequency at which all sources operate, we can deduce the values of the capacitors and inductors representing the complex coefficients. Of course these values depend on the chosen frequency. The DC-transformers turn into “normal” transformers; one has to compensate for their self-inductance. Instead of transformers one can take frequency-independent impedance converters.

## 5.6 Gradient Search

In contrast to Sec. 5.5.1, we assume again to have real random variables. We start with the problem given in Example 1 of Sec. 2.1 where we found the system of equations

$$\begin{cases} \frac{\partial}{\partial x_1} L = \frac{x_1}{\sigma_1^2} - \lambda & \stackrel{!}{=} 0 \\ \frac{\partial}{\partial x_2} L = \frac{x_2}{\sigma_2^2} - \lambda & \stackrel{!}{=} 0 \\ \frac{\partial}{\partial \lambda} L = y - x_1 + x_2 & \stackrel{!}{=} 0 \end{cases} \quad (5.22)$$

Minimizing a function  $f(\mathbf{x})$  using a gradient search type method proceeds as follows. One chooses a starting point  $\mathbf{x}^{(0)}$  and computes a new point by the update rule

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha \cdot \nabla_{\mathbf{x}} f(\mathbf{x}^{(k)}), \quad (5.23)$$

where the step-size  $\alpha$  has to be chosen appropriately. Instead of a discrete-time update, we can easily derive a continuous-time update formula

$$\frac{\partial}{\partial t} \mathbf{x}(t) = -\alpha' \cdot \nabla_{\mathbf{x}} f(\mathbf{x}(t)). \quad (5.24)$$

If the minimization problem includes constraints, then there is the following possibility. Choose an initial vector  $\mathbf{x}$  that fulfills all constraints and take an optimization procedure that guarantees that this constraints remain fulfilled. In terms of Eq. (5.22) this means that  $(x_1(t), x_2(t))$  should always fulfill  $y - x_1(t) - x_2(t) = 0$ .

Continuing the above example, we can generalize the equations in Eq. (5.22) to

$$\begin{cases} \frac{x_1(t)}{\sigma_1^2} - \lambda(t) & \stackrel{!}{=} -C_1 \frac{\partial}{\partial t} x_1(t) \\ \frac{x_2(t)}{\sigma_2^2} - \lambda(t) & \stackrel{!}{=} -C_2 \frac{\partial}{\partial t} x_2(t) \\ y - x_1(t) + x_2(t) & \stackrel{!}{=} 0 \end{cases} \quad (5.25)$$

and derive the electrical network shown in Fig. 5.11 where we introduced  $R_1 = \sigma_1^2$  and  $R_2 = \sigma_2^2$ . Rewriting Eq. (5.25) we get

$$\begin{pmatrix} C_1 & 0 \\ 0 & C_2 \\ 0 & 0 \end{pmatrix} \frac{\partial}{\partial t} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = - \begin{pmatrix} 1/R_1 & 0 & -1 \\ 0 & 1/R_2 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \lambda(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ y \end{pmatrix}. \quad (5.26)$$

As mentioned above, the initial voltages over the capacitors, which represents the initial guess, must sum to  $y$ .

Of course, starting from an electrical network like the one in Fig. 5.11, we can derive the underlying interpretation of gradient search in a minimization problem and relate that to a blockwise MAP estimation problem. Additional examples of circuits which could be analyzed in such a way are shown in Figs. 5.12, 5.13, and 5.14 (see also Fig. 9.4 of [4]).

---

<sup>6</sup>Note that there is no 1/2-factor in the exponent.



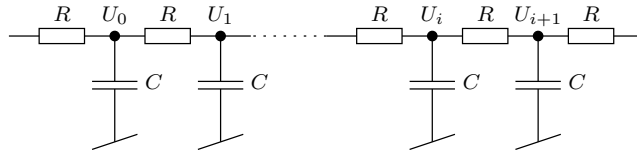


Figure 5.12: Chain of resistors and capacitors.

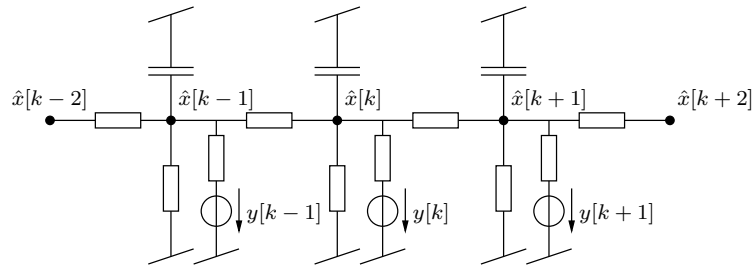


Figure 5.13: SKF with capacitors at estimated voltages.

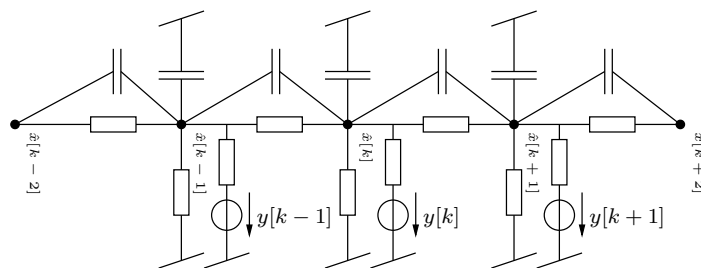


Figure 5.14: SKF with capacitors between any two estimated voltages and ground.

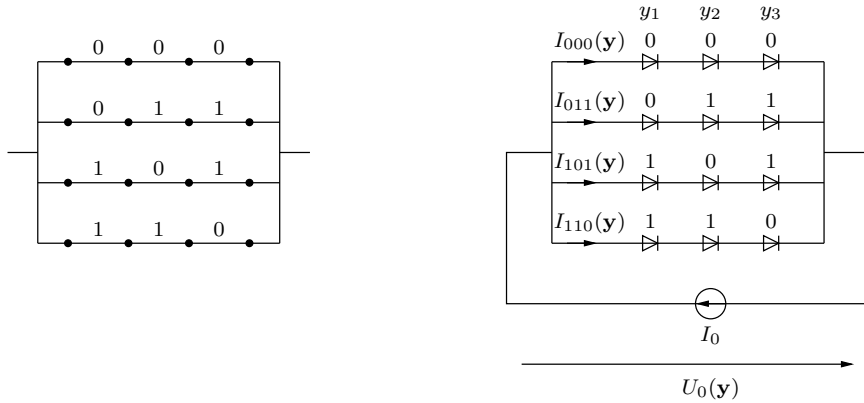


Figure 5.15: Left: canonical trellis of binary [3,2] linear code. Right: diode decoder derived from the canonical trellis of the binary [3,2] linear code.

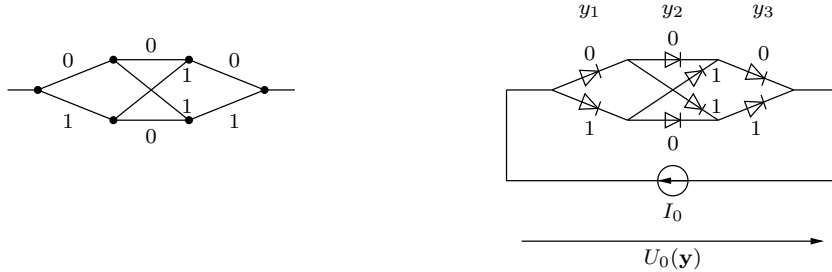


Figure 5.16: Left: minimal trellis of binary [3,2] linear code. Right: diode decoder derived from the minimal trellis of the binary [3,2] linear code.

## 5.7 Interpretation of the Diode-Decoder

Very much related to the approach taken in this report is the electrical network proposed by Davis and Loeliger [2] (see also the circuits for finding the shortest path in Dennis [1]). We start by considering a simple example in Sec. 5.7.1 and give then in Secs. 5.7.2 and 5.7.3 two possible analytical treatments. The idea is to find under what conditions the currents represent a-posteriori probabilities of the underlying problem.

### 5.7.1 A Simple Example

We start with a simple [3,2] binary linear code  $\mathcal{C}$  represented by the generator matrix

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \quad (5.27)$$

Its codewords are (0,0,0), (1,0,1), (0,1,1), and (1,1,0), respectively, which can be represented by the canonical trellis as in Fig. 5.15(left) or a minimal trellis as in Fig. 5.16(left). Figs. 5.15(right) and 5.16(right) give the corresponding diode decoder of this code [2] which uses diodes, switches and an ideal current source.

Let  $\mathbf{x}$  be a codeword, let  $\mathbf{y}$  be the received (hard-decision) word, and let  $d_H(\mathbf{x}, \mathbf{y})$  be their Hamming distance (i.e. the number of different symbols). A diode with a number means that if  $y_i$  equals this number then this diode is short-circuited (diode is “inactive”), and if  $y_i$  does not equal this number then this diode is not short-circuited (diode is “active”). ML decoding is the task of finding the codeword  $\mathbf{x}$  with the smallest Hamming distance to the received word  $\mathbf{y}$ . The current will search its way through the electrical network where there are the smallest numbers of diodes. ML decoding then corresponds to the task of finding where the largest current flows.

### 5.7.2 Canonical Trellis: The General Case

We assume that the diodes have the characteristic

$$I = I_s \cdot (\exp(U/U_T) - 1) \approx I_s \cdot \exp(U/U_T). \quad (5.28)$$

In general, we can consider an  $[n, k]$  binary linear code which has length  $n$  and dimension  $k$ . Let  $\mathbf{x}$  be a codeword, let  $\mathbf{y}$  be the received (hard-decision) word, and let  $d_H(\mathbf{x}, \mathbf{y})$  be their Hamming distance (i.e. the number of different symbols). We consider the diode decoder of the canonical trellis. The number of active diodes in branch  $\mathbf{x}$  is therefore  $d_H(\mathbf{x}, \mathbf{y})$ . The voltage  $U_0(\mathbf{y})$  is<sup>7</sup>

$$U_0(\mathbf{y}) = d_H(\mathbf{x}, \mathbf{y}) \cdot U_T \cdot \ln(I_{\mathbf{x}}(\mathbf{y})/I_s), \quad (5.29)$$

for each  $\mathbf{x} \in \mathcal{C}$ , from which we get

$$I_{\mathbf{x}}(\mathbf{y}) = I_s \cdot \exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x}, \mathbf{y})). \quad (5.30)$$

Using the Kirchhoff current law we get

$$I_0 = \sum_{\mathbf{x} \in \mathcal{C}} I_{\mathbf{x}}(\mathbf{y}) = I_s \cdot \sum_{\mathbf{x} \in \mathcal{C}} \exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x}, \mathbf{y})). \quad (5.31)$$

For a given  $I_0$ , this gives the relation from which the value of  $U_0(\mathbf{y})$  can be determined. For a given  $\mathbf{y}$ , the right-hand side is a monotonically increasing function in  $U_0(\mathbf{y})$  as it is the sum of monotonically increasing functions in  $U_0(\mathbf{y})$ ; the equation is thus uniquely solvable for  $U_0(\mathbf{y})$ . Note that  $U_0(\mathbf{y})$  is a function of all the different Hamming distances of each codeword to  $\mathbf{y}$ . As  $\mathcal{C}$  is a linear code it follows that  $U_0(\mathbf{y}) = U_0(\mathbf{y} + \mathbf{c})$  if  $\mathbf{c} \in \mathcal{C}$ .

We define the joint pmf

$$P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) := \frac{\exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x}, \mathbf{y}))}{2^n I_0/I_s} \cdot [\mathbf{x} \in \mathcal{C}]. \quad (5.32)$$

The fact  $\sum_{\mathbf{x}, \mathbf{y}} P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = 1$  follows from Eq. (5.31). From Eq. (5.31) follows also

$$P_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{C}} P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \frac{1}{2^n}. \quad (5.33)$$

---

<sup>7</sup>As  $U_0$  depends on the received word  $\mathbf{y}$  we write  $U_0 = U_0(\mathbf{y})$ .

For  $\mathbf{c} \in \mathcal{C}$  we have  $U_0(\mathbf{y}' + \mathbf{c}) = U_0(\mathbf{y}')$  and so we can observe that

$$P_{\mathbf{X}}(\mathbf{x} + \mathbf{c}) = \sum_{\mathbf{y}} P_{\mathbf{X}\mathbf{Y}}(\mathbf{x} + \mathbf{c}, \mathbf{y}) = \sum_{\mathbf{y}} \frac{\exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x} + \mathbf{c}, \mathbf{y}))}{2^n I_0/I_s} \cdot [\mathbf{x} + \mathbf{c} \in \mathcal{C}] \quad (5.34)$$

$$= \sum_{\mathbf{y}'} \frac{\exp(U_0(\mathbf{y}' + \mathbf{c})/U_T/d_H(\mathbf{x} + \mathbf{c}, \mathbf{y}' + \mathbf{c}))}{2^n I_0/I_s} \cdot [\mathbf{x} + \mathbf{c} \in \mathcal{C}] \quad (5.35)$$

$$= \sum_{\mathbf{y}'} \frac{\exp(U_0(\mathbf{y}')/U_T/d_H(\mathbf{x}, \mathbf{y}'))}{2^n I_0/I_s} \cdot [\mathbf{x} \in \mathcal{C}] \quad (5.36)$$

$$= \sum_{\mathbf{y}'} P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}') = P_{\mathbf{X}}(\mathbf{x}), \quad (5.37)$$

i.e.,

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|\mathcal{C}|} \cdot [\mathbf{x} \in \mathcal{C}] = \frac{1}{2^k} \cdot [\mathbf{x} \in \mathcal{C}]. \quad (5.38)$$

We get the conditional and a-posteriori probabilities

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{P_{\mathbf{X}}(\mathbf{x})} = \frac{\exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x}, \mathbf{y}))}{2^{n-k} I_0/I_s} \cdot [\mathbf{x} \in \mathcal{C}], \quad (5.39)$$

$$P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{P_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{P_{\mathbf{Y}}(\mathbf{y})} = \frac{\exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x}, \mathbf{y}))}{I_0/I_s} \cdot [\mathbf{x} \in \mathcal{C}]. \quad (5.40)$$

$$(5.41)$$

Therefore, for  $\mathbf{x} \in \mathcal{C}$ ,

$$\frac{I_{\mathbf{x}}(\mathbf{y})}{I_0} = \frac{I_s \cdot \exp(U_0(\mathbf{y})/U_T/d_H(\mathbf{x}, \mathbf{y}))}{I_0} = P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = 2^{n-k} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}). \quad (5.42)$$

Measuring the ratio  $I_{\mathbf{x}}(\mathbf{y})/I_0$  gives both  $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$  and  $P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$  from which we can make a blockwise MAP or a blockwise ML decision. These are of course the same, as can also be seen from the fact that  $P_{\mathbf{X}}(\mathbf{x})$  is uniform over the codewords.

For  $\mathbf{c} \in \mathcal{C}$  one can prove that  $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} + \mathbf{c}|\mathbf{x} + \mathbf{c}) = P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$  and  $P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} + \mathbf{c}|\mathbf{y} + \mathbf{c}) = P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$  in a similar fashion as we proved  $P_{\mathbf{X}}(\mathbf{x} + \mathbf{c}) = P_{\mathbf{X}}(\mathbf{x})$ .

It is obvious that at most one  $\mathbf{x}$  can have  $d_H(\mathbf{x}, \mathbf{y}) = 0$ , in which case  $I_{\mathbf{x}}(\mathbf{y}) = I_0$  and  $I_{\mathbf{x}'} = 0$  for all  $\mathbf{x}' \in \mathcal{C}$ ,  $\mathbf{x}' \neq \mathbf{x}$ .

### 5.7.3 General Trellis: The General Case

It is straightforward to give the FFGs representing the diode decoders in Figs. 5.15 and 5.16, see Figs. 5.17 and 5.18. To go from the FFG to the electrical network one can use the rules of Ch. 3, but this time the Kirchhoff current laws have been interpreted as Kirchhoff voltage laws and vice-versa.<sup>8</sup> Thereby we introduced new function nodes  $f^{(0)}(.,.)$  and  $f^{(1)}(.,.)$  in Fig. 5.17(right). We explain its meaning for the upper case, the lower follows by analogy. If

<sup>8</sup>So, roughly speaking, an equality-function node corresponds to serial concatenation, whereas a summation-function node corresponds to parallel concatenation. Moreover, the currents give the estimated values.

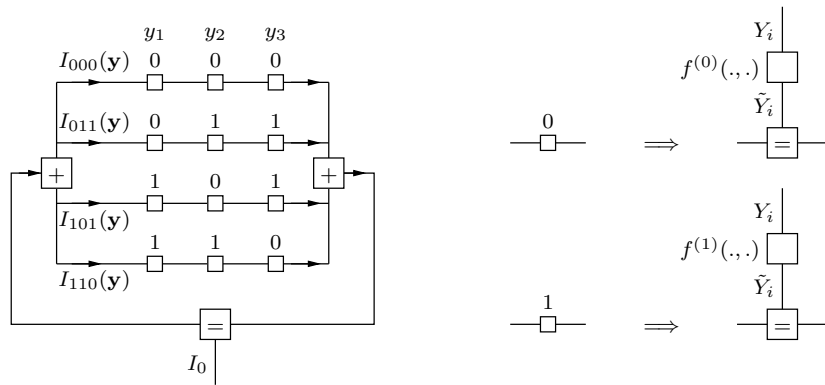


Figure 5.17: Left: FFG interpretation of a diode decoder of a canonical trellis of a binary [3, 2] linear code. Right: Meaning of function nodes used in the left part (see also text).

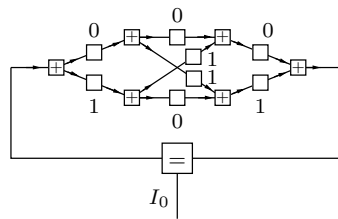


Figure 5.18: FFG interpretation of a diode decode of a minimal trellis of a binary [3, 2] linear code. (The same function nodes were used as defined in Fig. 5.17(right).)

$y_i$  is 0, then the distribution  $f^{(0)}(0, \tilde{y}_i)$  of  $\tilde{Y}_i$  is a “short circuit” distribution, whereas if  $y_i$  is 1 then the distribution  $f^{(0)}(1, \tilde{y}_i)$  of  $\tilde{Y}_i$  is the “diode distribution” (as in Tab. G.3, but with current and voltage interchanged).

The idea is that the currents represent some estimate of the a-posteriori probabilities. Of course, instead of the above distributions one can take also others, and one can consider its dual electrical network (see also Ch. 7).

## Chapter 6

# Primal-Dual FFGs and Multiports

This chapter introduces first a slightly reformulated objective function of the primal problem compared with Ch. 3 and looks then at the dual problem. The second part reformulates the results of Ch. 3 using the concept of multiports and then considers the implications of primal and dual problem to electrical networks. Roughly speaking, we show that when the voltages solve the primal problem, then the currents solve the dual problem and vice-versa. An important role will be played by Tellegen's Theorem. At the end we mention also other dualities. We will use the nomenclature of variables and functions as defined at the beginning of App. C.

### 6.1 The Primal and the Dual Problem

#### 6.1.1 The Primal Problem

We consider the same maximization/minimization problem as in Ch. 3, but we write the global distribution slightly differently so that the proofs are easier. The basic idea is to replace Dirac-delta distributions by Gaussian distributions. But to simplify matters, we do it with a little twist compared to Sec. 1.6. Assume that we are optimizing over  $\mathbf{x} = (x_1, x_2, x_3)$ , i.e. a three-dimensional space. We replace now distribution like  $\delta(x_1)$  by

$$\exp\left(-\frac{1}{2\beta}x_1^2\right) \cdot \exp\left(-\frac{\beta}{2}x_2^2\right) \cdot \exp\left(-\frac{\beta}{2}x_3^2\right), \quad (6.1)$$

which up to scaling factors is a Gaussian distribution with variance  $\beta$  in direction  $x_1$  and variance  $1/\beta$  in directions  $x_2$  and  $x_3$ .

Accordingly, a distribution like  $\delta(x_1 - x_2)$  would be replaced by a Gaussian distribution (without prefactor) having variance  $\beta$  in direction  $(1, -1, 0)$  and variance  $1/\beta$  in the directions orthogonal to  $(1, -1, 0)$ .

In the limit  $\beta \rightarrow 0$  of this procedure, the optimizing vectors in the optimization problems we are interested in are unchanged compared to the optimizing vectors in the earlier formulations. We apply the following modifications to an FFG.

- All constraint function nodes (equality-function nodes and summation-function nodes) originally expressed using Dirac-delta distributions are replaced by scaled Gaussian distributions as just explained.

- Half-edges  $X_i$  where a measurement  $X_i = x_i^{(M)}$  is available are extended to full-edges by adding a function node  $\delta(\cdot - x_i^{(M)})$  which is then also expressed by a scaled Gaussian distribution as explained above.
- Each edge  $X_i$  gets replaced by  $X'_i$  and  $X''_i$ . (Firstly, it is immaterial which end is primed and which end is double-primed; secondly it is immaterial which arguments of a function are primed and which are double-primed.)
- As shown at the beginning of App. C, we introduce  $\phi_t(x_t) := -\ln f_t(x_t)$  for each leaf-node function  $f_t(x_t)$  (we assume  $\phi_t(\cdot)$  to be convex).

6.1(a and b) show an example of a modified FFG.<sup>1</sup> With the above reformulation, the optimization problem of Ch. 3 turns into the following minimization problem

$$\min_{\mathbf{x}: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}) \quad (6.2)$$

with

$$\mathbf{x} = (x'_1 \quad x''_1 \quad x'_2 \quad x''_2 \quad \cdots \quad x'_N \quad x''_N)^T \quad (6.3)$$

$$\mathbf{A} = \begin{pmatrix} +1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & +1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & +1 & -1 \end{pmatrix} \quad (6.4)$$

$$\mathbf{b} = (0 \quad 0 \quad \cdots \quad 0)^T. \quad (6.5)$$

The objective function thereby is

$$\phi(\mathbf{x}) := \sum_{t \in \mathcal{F}} \phi_t(\mathbf{x}_t), \quad (6.6)$$

where the summation is over all function nodes  $\mathcal{F}$ : for a  $t \in \mathcal{F}$ ,  $\phi_t(\cdot)$  is the function with label  $t$ ,  $\mathbf{x}_t$  are its arguments. Introducing the primed and double-primed edge labels we have achieved that the arguments to the different function nodes are all different (i.e. they do not overlap). But the constraint  $\mathbf{Ax} = \mathbf{b}$  takes care of the fact that the primed and double-primed edge-labels should be equal. Actually,  $\phi(\mathbf{x})$  is also a function of  $\beta$ , but we do not show it explicitly (we are always only interested in the limit  $\beta \rightarrow 0$ ).

### 6.1.2 The Dual Problem

With the preparations done in the last subsection, it is an easy matter now to formulate the dual problem. By Sec. C.2 we have (note that  $\mathbf{b} = \mathbf{0}$ )

$$\min_{\mathbf{x}: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}) = \max_{\boldsymbol{\xi}, \boldsymbol{\gamma}: \mathbf{A}^T \boldsymbol{\gamma} = \boldsymbol{\xi}} \boldsymbol{\gamma}^T \mathbf{b} - \phi^*(\boldsymbol{\xi}) = \max_{\boldsymbol{\xi}, \boldsymbol{\gamma}: \mathbf{A}^T \boldsymbol{\gamma} = \boldsymbol{\xi}} -\phi^*(\boldsymbol{\xi}) = \max_{\boldsymbol{\xi}', \boldsymbol{\xi}'', \boldsymbol{\xi}' = -\boldsymbol{\xi}''} -\phi^*(\boldsymbol{\xi}', \boldsymbol{\xi}''), \quad (6.7)$$

<sup>1</sup>Note the slight difference to Example 1 in Sec. 2.1: here we have  $X_3 = -(X_1 + X_2)$ , i.e.  $X_1 + X_2 + X_3 = 0$ .



with

$$\boldsymbol{\xi} = (\xi'_1 \ \xi''_1 \ \xi'_2 \ \xi''_2 \ \cdots \ \xi'_N \ \xi''_N)^T, \quad (6.8)$$

$$\boldsymbol{\xi}' = (\xi'_1 \ \xi'_2 \ \cdots \ \xi'_N)^T, \quad (6.9)$$

$$\boldsymbol{\xi}'' = (\xi''_1 \ \xi''_2 \ \cdots \ \xi''_N)^T, \quad (6.10)$$

$$\boldsymbol{\gamma} = (\gamma_1 \ \gamma_2 \ \cdots \ \gamma_N)^T \quad (6.11)$$

We mentioned at the end of Sec. 6.1.1 that the arguments of the different function nodes did not overlap, therefore the conjugate function must have the form

$$\phi^*(\boldsymbol{\xi}) = \sum_{t \in \mathcal{F}} \phi_t^*(\boldsymbol{\xi}_t), \quad (6.12)$$

i.e., we can conjugate each summand separately.

- Let  $\phi_t(\cdot) = -\ln f_t(\cdot)$  be the objective function of a leaf-function node. The conjugate  $\phi_t^*(\cdot)$  leads then to the dual generalized distribution  $f_t^*(\cdot) = \exp(-\phi_t^*(\cdot))$ .
- The dualization of constraint functions nodes (those where Dirac-delta distributions have been replaced by jointly Gaussian densities) is shown in Sec. C.5.

We finish this subsection by relating the optimized value of the primal and the optimized value of the dual problem. From Eq. (6.7) we have<sup>2</sup>

$$0 = \left( \min_{\mathbf{x}: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}) \right) - \left( \max_{\boldsymbol{\xi}, \boldsymbol{\gamma}: \mathbf{A}^T \boldsymbol{\gamma}=\boldsymbol{\xi}} -\phi^*(\boldsymbol{\xi}) \right) = \left( \min_{\mathbf{x}: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}) \right) + \left( \min_{\boldsymbol{\xi}, \boldsymbol{\gamma}: \mathbf{A}^T \boldsymbol{\gamma}=\boldsymbol{\xi}} \phi^*(\boldsymbol{\xi}) \right) \quad (6.13)$$

$$= \min_{\substack{\mathbf{x}: \mathbf{Ax}=\mathbf{b} \\ \boldsymbol{\xi}, \boldsymbol{\gamma}: \mathbf{A}^T \boldsymbol{\gamma}=\boldsymbol{\xi}}} \phi(\mathbf{x}) + \phi^*(\boldsymbol{\xi}), \quad (6.14)$$

i.e., for any  $\mathbf{x}$ ,  $\boldsymbol{\xi}$ ,  $\boldsymbol{\gamma}$  satisfying  $\mathbf{Ax} = \mathbf{b}$  and  $\mathbf{A}^T \boldsymbol{\gamma} = \boldsymbol{\xi}$  we have

$$\phi(\mathbf{x}) + \phi^*(\boldsymbol{\xi}) \geq 0. \quad (6.15)$$

Let  $(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\gamma})$  minimize the expression in Eq. (6.14), then  $\boldsymbol{\xi}$  must be the gradient of  $\phi(\cdot)$  at the point  $\mathbf{x}$ , and  $\mathbf{x}$  must be the gradient of  $\phi^*(\cdot)$  at the point  $\boldsymbol{\xi}$ . (This follows from the derivations in Sec. C.2). More on this topic can be found in Sec. 6.2.3.

### 6.1.3 Primal FFG, Dual FFG, and Primal-Dual FFG

The aim of this subsection is to consider the similarities of the FFGs related to the primal and the dual problem, respectively. Fig. 6.1(a) shows the FFG related to a primal problem, and Fig. 6.1(b) a slightly modified version as defined in Sec. 6.1.1. We call such a graph PFFG for primal FFG.

It is possible to define a dual FFG (called DFFG) that represents a generalized distribution which leads to the conjugate function as in Eq. (6.12). Because of the structural similarity of  $\phi(\cdot)$  in Eq. (6.6) and  $\phi^*(\cdot)$  in Eq. (6.12), it is not difficult to see that the PFFG and the DFFG must have the same topology but with the following changes.

<sup>2</sup>Such an equality was also given on p. 34 of [1] for the quadratic minimization problem.

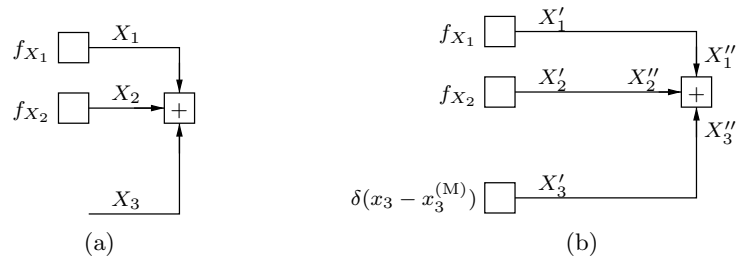


Figure 6.1: Primal FFG (PFFG). Each edge label gets replaced by two edge labels: one primed and the other double-primed.

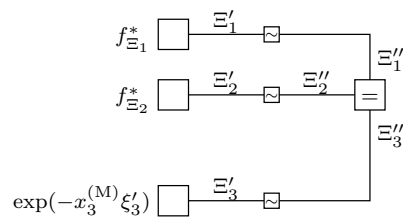


Figure 6.2: Dual FFG (DFFG).

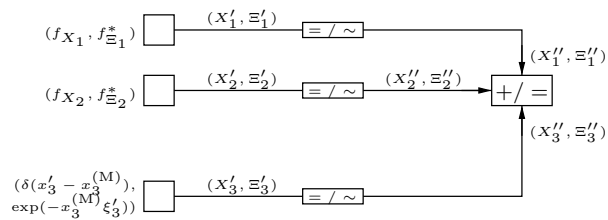


Figure 6.3: Primal-Dual FFG (PDFFG).

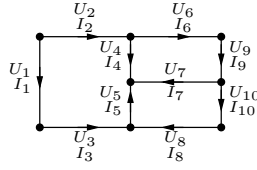


Figure 6.4: Illustration of Tellegen's Theorem: graph representing the topology of an electrical network. The voltages (between nodes) and currents (on the branches) are measured in the same direction.

- A leaf function  $f_t(\cdot)$  (which is a generalized distribution function) is replaced by its dual generalized distribution function  $f^*(\cdot)$  where  $\phi_t(\cdot) = -\ln f_t(\cdot)$  and  $\phi_t^*(\cdot) = -\ln f_t^*(\cdot)$ .
- Constraint function nodes are replaced by their dual constraint function nodes (see Sec. C.5).
- Measurement function nodes  $\delta(x_t - x_t^{(M)})$  (expressed by a scaled Gaussian distribution as in Sec. 6.1.1) are replaced by the function  $\exp(-x_t^{(M)}\xi_t)$ .
- Each variable edge  $X_i$  (with primal variables  $X_i'$  and  $X_i''$ ) has to be replaced by two edges labelled by the dual variables  $\Xi_i'$  and  $\Xi_i''$ , respectively, connected by an inverter function node in order to fulfill  $\Xi' = -\Xi''$ . (The symbol of an inverter-function node can be seen e.g. in Fig. 6.2; from a function-node point of view it is equivalent to a summation-function node with both arrows pointing inwards.)

Fig. 6.2 shows the DFFG of the PFFG in Fig. 6.1(b). Because of the topological equivalence, we can try to unify the PFFG and the DFFG to a primal-dual FFG (PDFFG) as in Fig. 6.3:

- Each edge label has two components: one from the primal problem and one from the dual problem.
- Each function node bears two labels: one for the primal function node and one for the dual function node.
- The inverter-function nodes of the DFFG are expanded to a combination of an equal-function node (for the PFFG) and an inverter-function node (for the DFFG).

## 6.2 Tellegen's Theorem and Multiports

We switch now from primal and dual problems to electrical networks. The results of Ch. 3 can be cast using the concept of multiports.

### 6.2.1 Tellegen's Theorem

**Theorem 1 (Tellegen's Theorem)** Consider a graph representing the topology of an electrical network (see e.g. Fig. 6.4). Let  $\{U_i\}$  be an assignment of branch voltages that satisfies

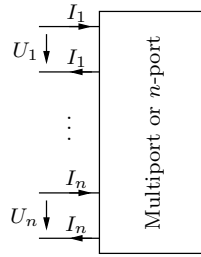


Figure 6.5:  $n$ -port: multipoint with  $n$  ports. By definition of a multipoint, the same current going in at one terminal of a port goes out at the other terminal of the same port.

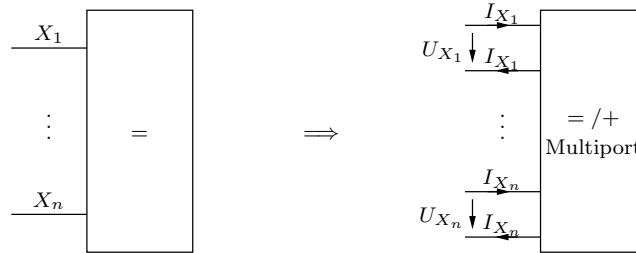


Figure 6.6: Equality-function node and the voltage-equality (current-summation) multipoint.

the Kirchhoff voltage law, and let  $\{I_i\}$  be an assignment of branch currents that satisfies the Kirchhoff current law. Then

$$\sum_i U_i I_i = 0. \quad (6.16)$$

(Note that we did not assume any connection between the voltages  $U_i$  and the currents  $I_i$ .)

*Proof:* For a proof, see e.g. Ch. 12.3 of [16].  $\square$

One application of Tellegen's theorem is to realize that all the power that is created by active elements is exactly used by the passive elements and the overall power sum is zero.

## 6.2.2 Multiports

We start by giving the definition of a multipoint, or  $n$ -port.<sup>3</sup>

**Definition 2 (Multipoint,  $n$ -port)** We define a multipoint ( $n$ -port) (see Fig. 6.5) as an abstract electrical networks with  $n$  pairs of terminals where for each pair the same current going in at one terminal goes out at the other terminal. The voltage between the two terminals (positive to negative terminal) of a port is denoted by  $U_i$  and the current flowing in at the positive terminal by  $I_i$ . It is required that there must be exactly  $n$  linearly independent equations among all voltages  $\{U_i\}$  and currents  $\{I_i\}$ .

<sup>3</sup>More information can be found e.g. in [17].

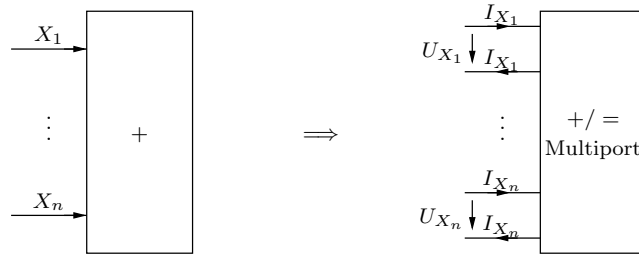


Figure 6.7: Summation-function node and the voltage-summation (current-equality) multiport.

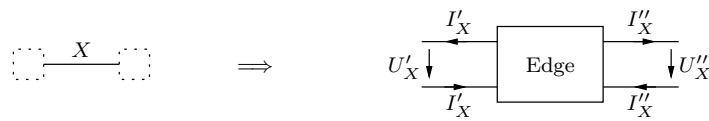


Figure 6.8: Edge and edge multiport. (Note that the orientation of the currents has been chosen so that they fit the current convention of the other multiports it is connected to.)

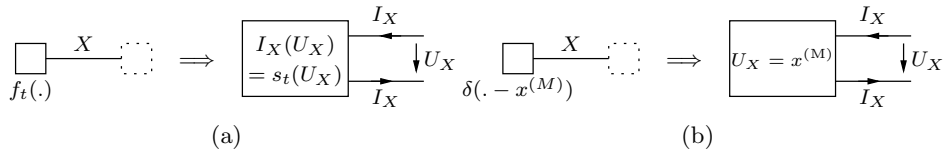


Figure 6.9: (a) Function node and function multiport. (b) As a special case, the measurement function node and the measurement function multiport.

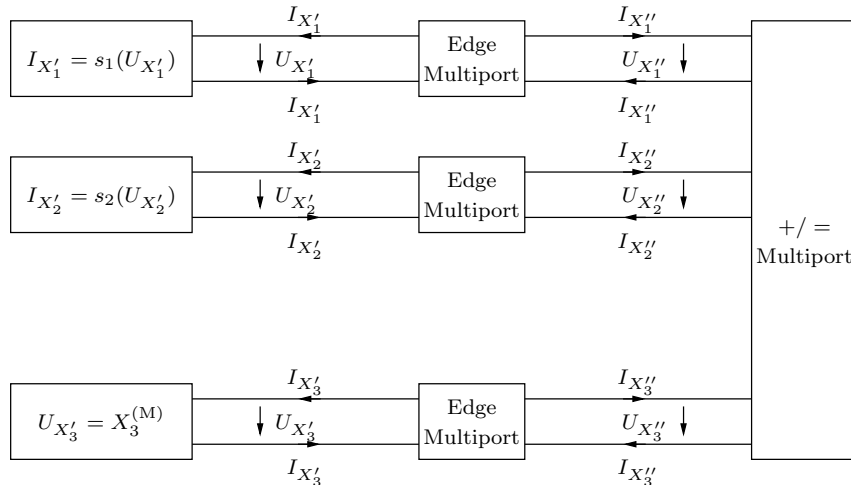


Figure 6.10: Multiport representation with a voltage-summation (current-equality) multiport of the electrical network derived from the PFFG in Fig. 6.1.

The power consumption of a multiport can be completely characterized by knowing all input voltages and input currents.

**Lemma 3** *The power consumption of a multiport is*

$$\sum_{i=1}^n U_i I_i = P. \quad (6.17)$$

*Especially, if the multiport is neutral (i.e.  $P = 0$ ) then*

$$\sum_{i=1}^n U_i I_i = 0. \quad (6.18)$$

*Proof:* We connect between each pair of terminal elements with arbitrary current-voltage characteristic that do not contradict the internal relations of the multiport. Let  $U'_i$  and  $I'_i$  be the branch voltages and currents in the whole electrical network. Using Tellegen's Theorem (Th. 1) in the first equality we get

$$0 = \sum_{i: \text{ all branches}} U'_i I'_i = \underbrace{\sum_{i: \text{ inside multiport}} U'_i I'_i}_{=P} + \underbrace{\sum_{i: \text{ outside multiport}} U'_i I'_i}_{=\sum_{i=1}^n U_i (-I_i)}, \quad (6.19)$$

from which the desired equality follows.  $\square$

The electrical networks derived in Ch. 3 can be reformulated in the context of multiports. Figs. 6.6, 6.7, 6.8, and 6.9(a and b) show the multiport for an equality-function node, a summation-function node, an edge, a general function node and a measurement function node. Note that in each case it is guaranteed that the current flowing in at one terminal of a port is the same as the current flowing out of the other terminal (this follows from the derivations in Ch. 3), therefore we can use the concept of multiports. Fig. 6.10 gives the multiport representation of the electrical network derived from the PFFG in Fig. 6.1. As can be seen from the explicit partial electrical networks in Ch. 3, these multiports have the following properties.

- The equality multiports and summation multiports are both neutral, so that Eq. (6.18) holds for both of them. This means that the voltage vector and the current vector are orthogonal to each other. As a consequence for the equality multiport, the sum of all incoming currents must be zero, and as a consequence for the summation multiport, all incoming currents must be equal.<sup>4</sup>
- The edge multiport is also neutral. The special feature is that the voltages  $U'_X$  and  $U''_X$  are equal whereas the currents  $I'_X$  and  $I''_X$  are inverses of each other, i.e.,  $I'_X = -I''_X$ . (The direction of the currents have been defined so that they match the directions of the currents of the function multiports the edge is connected to.)
- General function nodes (and measurement function nodes) are usually active or passive (i.e. not neutral).

---

<sup>4</sup>This theoretically derived last result indeed happens as can be seen from the partial electrical networks in Ch. 3. .

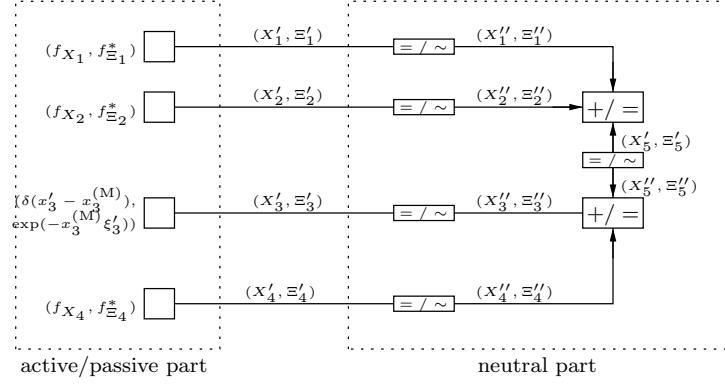


Figure 6.11: The active/passive and the neutral parts of a PDFFG.

We would like to argue now that the PDFFG in Fig. 6.3 and the electrical network in Fig. 6.10 are “equivalent”. We only have to see that

- $x'_i$  corresponds to  $U_{X'_i}$ , and  $\xi'_i$  corresponds to  $I_{X'_i}$  (same for  $x''_i$ ,  $\xi''_i$ ,  $U_{X''_i}$ , and  $I_{X''_i}$ ).
- $\boxed{= / \sim}$ -function nodes correspond to edge multiports:  $x'_i = x''_i$  corresponds to  $U_{X'_i} = U_{X''_i}$ , and  $\xi'_i = -\xi''_i$  corresponds to  $I_{X'_i} = -I_{X''_i}$ .
- $\boxed{= / +}$ -function nodes correspond to equal multiports. In an example with  $n = 3$  we would have that  $x''_1 = x''_2 = x''_3$  corresponds to  $U_{X''_1} = U_{X''_2} = U_{X''_3}$ , and that  $\xi''_1 + \xi''_2 + \xi''_3 = 0$  corresponds to  $I_{X''_1} + I_{X''_2} + I_{X''_3} = 0$ .
- $\boxed{+ / =}$ -function nodes correspond to summation multiports. Using the example in Fig. 6.10,  $x''_1 + x''_2 + x''_3 = 0$  corresponds to  $U_{X''_1} + U_{X''_2} + U_{X''_3} = 0$ , and  $\xi''_1 = \xi''_2 = \xi''_3$  corresponds to  $I_{X''_1} = I_{X''_2} = I_{X''_3}$ .
- Leaf-function nodes correspond to 1-ports. The score function  $\xi'_i = s(x'_i)$  corresponds to the current-voltage characteristic  $I_{X'_i} = I_{X'_i}(U_{X'_i}) = s(U_{X'_i})$ , and the dual score function  $x'_i = s^*(\xi'_i)$  corresponds to the voltage-current characteristic  $U_{X'_i} = U_{X'_i}(I_{X'_i}) = s(I_{X'_i})$ .

### 6.2.3 Active/Passive Part versus Neutral Part

The nodes of an FFG can be grouped into an “active/passive part” and a “neutral part”, as shown for a specific example in Fig. 6.11. The idea behind this classification is the following.

- The function nodes in the active/passive part will be represented by active/passive multiports in an electrical network.
- The function nodes in the neutral part will be represented by neutral multiports in an electrical network.

As the electrical network has the same topology, the same grouping is possible. We would like to make the following comments about the neutral part.

- The dual behaviors studied e.g. in [7] correspond to the neutral part where duality is defined on the variables that are between the two parts. (In Fig. 6.11 this means  $x'_1\xi'_1 + x'_2\xi'_2 + x'_3\xi'_3 + x'_4\xi'_4 = 0$ , see also Sec. D.3.)
- As long as the primal edge variables are a valid configuration and the dual edge variables are a valid configuration, they are completely independent (compare this with the statement for the active/passive part).
- The previous statement means for the electrical networks that voltages and currents are “treated separately” in the sense that there are no component laws that tie voltages and currents somehow together.

We would like to make the following comments about the active/passive part.

- When the primal and dual edge variables assume the value corresponding to the block-wise MAP estimate, they are tied together through the score function.
- The last statement means for electrical networks that in the active/passive part there are component laws that tie voltages and currents somehow together.

We conclude with some power considerations. We assume to have the multiport representation of an electrical network derived from an FFG. Then we have equality in Lemma 9 of Sec. C.6 for all primal and dual variables. Combined with Lemma 3 of Sec. 6.2.2 this means that the sum of the power consumption of all parts equals (the summation over all  $i$  just means over all ports of all multiports and the summation over all  $j$  means over all edges)

$$P_{\text{total}} = \sum_i U_i I_i = \sum_j (x'_j \xi'_j + x''_j \xi''_j) = \sum_{t \in \mathcal{F}} (\phi_t(\mathbf{x}_t) + \phi_t^*(\boldsymbol{\xi}_t)) \quad (6.20)$$

$$= \left( \sum_{t \in \mathcal{F}} \phi_t(\mathbf{x}_t) \right) + \left( \sum_{t \in \mathcal{F}} \phi_t^*(\boldsymbol{\xi}_t) \right) = \phi(\mathbf{x}) + \phi^*(\boldsymbol{\xi}). \quad (6.21)$$

But from Tellegen's Theorem we know that  $P_{\text{total}} = 0$ , so that  $\phi(\mathbf{x}) + \phi^*(\boldsymbol{\xi}) \geq 0$  with equality when minimized over  $\mathbf{x}$  and  $\boldsymbol{\xi}$ . This confirms the results at the end of Sec. 6.1.2. Moreover, if at least some components of  $\mathbf{x}$  and  $\boldsymbol{\xi}$  are not tied together by the score function, then by Lemma 9 of Sec. C.6 we have  $\phi(\mathbf{x}) + \phi^*(\boldsymbol{\xi}) > 0$ .

#### 6.2.4 Dualizing Parts of an Electrical Network

In this section we present some results which are very akin of the statements in Sec. D.3. It can be shown (we give no proof here) that the two electrical networks in Fig. 6.12 are “from outside” electrically the same. The gyrators essentially “exchange voltage and current”. In terms of messages this can be stated as the gyrator switches the voltage and the current axes. Note that gyrators are neutral multiports.

Assuming the rightness of the network equality in Fig. 6.12, the two electrical networks in Fig. 6.13 can be shown to be equal “from outside”. First we dualize each summation-function multiport separately by the procedure in Fig. 6.12 (the intermediary result is shown in Fig. 6.14). Secondly, we have to realize that a serial concatenation of a gyrator, an edge multiport, and a gyrator (the gyrators have different directions) are equivalent to a  $1 : -1$  DC-transformer (we called it dual edge multiport as it is the dual of the edge multiport).

The above procedure not only works for electrical networks stemming from FFGs which are trees but also for electrical networks stemming from loopy FFGs.



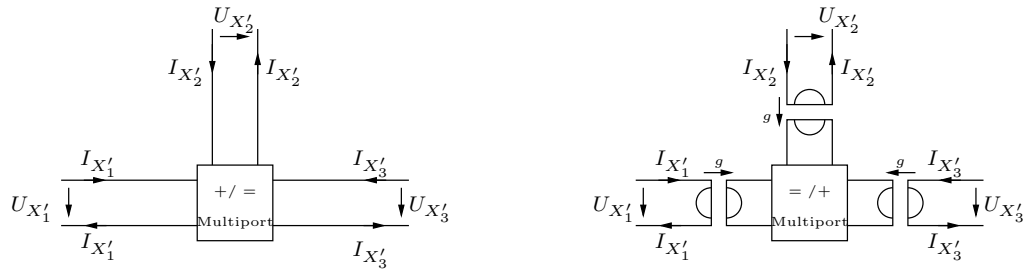


Figure 6.12: The left and the right partial electrical networks are the same “from outside”. Left: voltage-summation current-equality multiport. Right: voltage-equality current-summation multiport and gyrators.

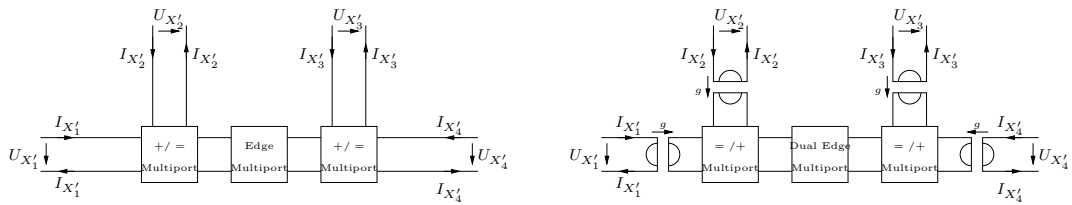


Figure 6.13: The left and the right partial electrical networks are the same “from outside”. Left: two voltage-summation current-equality multiports and an edge multiport. Right: two voltage-equality current-summation multiports, a dual edge multiport, and gyrators.

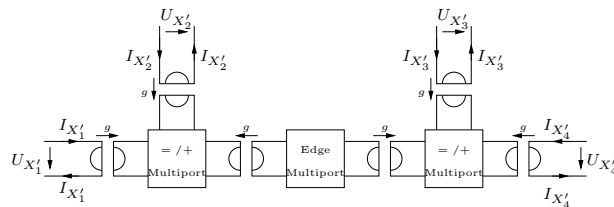


Figure 6.14: Intermediary step between the left and right part of Fig. 6.13.

## 6.3 Other Types of Dualities

Besides Lagrange Duality as considered in this chapter there are also other types of dualities. We quickly discuss the different possibilities, which are sometimes not completely unrelated to Lagrange duality.

### 6.3.1 Current-Voltage Duality

The equations we interpreted as Kirchhoff voltage law in Ch. 3 can also be interpreted as Kirchhoff current laws and vice-versa; the current-voltage characteristics turn into voltage-current-characteristics and vice-versa. The node-potential method is replaced by the loop-currents method.

### 6.3.2 Fourier Duality of FFGs

One of Forney's motivation to introduce normal factor graphs [7] was the reason that in the context of Fourier duality one can easily dualize an FFG that consists only of constraint function nodes (as e.g. in coding theory). Topologically it looks the same, but the constraint function nodes are replaced by their dual constraint function nodes, and state variable edges have to be split into two parts and connected by a negator function node (for details, see [7] and App. D). This type of constraint function node duality also shows up as part of the Lagrange duality as considered in this chapter and in App. C.

In the context of Fourier duality, general FFGs turn into convolution FFGs as was shown in [18].

### 6.3.3 Planar Graph Duality, Planar Electrical Network Duality, and Matroid Duality

As explained in App. E, in the same manner as planar graphs can be dualized, one can dualize planar electrical networks. Consider an FFG which is a tree and which models a jointly Gaussian distribution. We can derive two different electrical networks. The first is by choosing the edge variables to be voltages as in Ch. 3 and derive the electrical network accordingly. The other is to choose the edge variables to be currents and derive the corresponding electrical network. These two electrical networks are dual in the above sense of electrical network duality.

Recski [17] calls the dual electrical networks (as defined e.g. in App. E) "inverse" electrical networks because he reserved the term dual electrical network for a concept related to matroidal duality (see Sec. 10.1 of [17]). This dualization concept focuses on multiports: it turns out that the Lagrange dual of a subspace constraint function nodes is also the dual in this matroidal sense (independent if there is a planar representation or not).

## Chapter 7

# Conclusions and Outlook

The main aim of this report was to show that there is topologically a one-to-one connection between FFGs and electrical networks solving a blockwise MAP problem. Considering different types of dualities proved also to be a source of inspiration. We would like to conclude with a list of open questions.

- We considered only convex problems. But nothing stops us from using probability densities whose negative exponent is not convex and whose score function therefore is not a monotonically increasing function anymore. What solution will the electrical networks find?
- Can the result in Sec. 5.3 about measuring the error covariance matrix also be obtained by modifying slightly the underlying factor graph?
- The electrical networks as derived in this report give the correct blockwise MAP estimates independent of the fact if the FFG is a tree or not and as long as the maximization problem is convex. So, nature is able to find the correct solutions in electrical networks. On the other hand, the MPA does not always give the correct estimate back (see e.g. [13]). The question is if one can learn something from nature, i.e., can one somehow mimick its behavior?
- As an interesting example to the previous point we propose to study the tail-biting (or cyclic) Kalman filter of length  $n$ : after  $n$  sections the new state is not  $\mathbf{x}[n + 1]$ , but  $\mathbf{x}[1]$  again (the relationship between the usual Kalman filter and the tail-biting Kalman filter is the same as between a trellis and a tail-biting trellis in coding theory). What can be learned from this example?
- A first approach was taken to represent the diode decoder in the present framework, and we had a quick first look at some time-dependent circuits. More work has to be done in both directions.



# Appendix A

## Some Facts from Estimation Theory

Let  $\mathbf{X}$  be an  $n$ -dimensional random vector and let  $\mathbf{Y}$  be an  $m$ -dimensional random vector. We assume to know the joint density function  $p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ . Based on a measurement  $\mathbf{Y} = \mathbf{y}$  one wants to give an estimate  $\hat{\mathbf{X}} = \mathbf{g}(\mathbf{Y})$  on  $\mathbf{X}$ . We call  $\tilde{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{X}}$  the estimation error.  $E[\tilde{\mathbf{X}}]$  is the bias of the estimator and  $\text{Var}[\tilde{\mathbf{X}}]$  is the error covariance matrix of the estimator. An estimator which fulfills  $E[\tilde{\mathbf{X}}] = \mathbf{0}$ , is called bias-free.<sup>1</sup>

### A.1 A First Family of Estimators

#### A.1.1 Conditional Expectation Estimator

This estimator gives the conditional expectation as estimate.

$$\hat{\mathbf{x}}^E = \mathbf{g}^E(\mathbf{y}) := E[\mathbf{X}|\mathbf{Y} = \mathbf{y}]. \quad (\text{A.1})$$

#### A.1.2 Quadratic Cost Function

Let  $\mathbf{Q}$  be a symmetric positive-definite matrix and let the cost function be

$$C^{\mathbf{S},\mathbf{Q}}(\tilde{\mathbf{x}}) := \tilde{\mathbf{x}}^T \mathbf{Q} \tilde{\mathbf{x}}. \quad (\text{A.2})$$

Let

$$\hat{\mathbf{x}}^{\mathbf{S},\mathbf{Q}} = \mathbf{g}^{\mathbf{S},\mathbf{Q}}(\mathbf{y}). \quad (\text{A.3})$$

be the estimator  $\mathbf{g}$  which minimizes the expected cost

$$E[C^{\mathbf{S},\mathbf{Q}}(\mathbf{X} - \mathbf{g}(\mathbf{Y}))]. \quad (\text{A.4})$$

#### A.1.3 Estimator with Smallest Expected Square Error

The estimator with smallest expected square error is the estimator with cost matrix  $\mathbf{Q} = \mathbb{1}$ ,

$$\hat{\mathbf{x}}^{\mathbf{S}} = \mathbf{g}^{\mathbf{S}}(\mathbf{y}) = \mathbf{g}^{\mathbf{S},\mathbb{1}}(\mathbf{y}). \quad (\text{A.5})$$

---

<sup>1</sup>Sometimes the stronger  $E[\tilde{\mathbf{X}}|\mathbf{Y} = \mathbf{y}] = \mathbf{0}$  for each  $\mathbf{y}$  is required for bias-freeness.

### A.1.4 Bias-Free Minimum-Variance Estimator

The bias-free minimum-variance estimate

$$\hat{\mathbf{x}}^{\text{BFMV}} = g^{\text{BFMV}}(\mathbf{y}), \quad (\text{A.6})$$

is the estimate where  $\mathbf{g}^{\text{BFMV}}$  is the bias-free estimator  $\mathbf{g}$  which minimizes  $\sum_{i=1}^n \text{Var}[X_i - g_i(\mathbf{Y})]$ .

### A.1.5 Properties of the First Family of Estimators

All the estimators from the first family are equal

$$\mathbf{g}^{\text{E}} = \mathbf{g}^{\text{S,Q}} = \mathbf{g}^{\text{S}} = \mathbf{g}^{\text{BFMV}} \quad (\text{independent of } \mathbf{Q}), \quad (\text{A.7})$$

and are bias-free

$$\mathbb{E}[\tilde{\mathbf{X}}] = \mathbf{0}, \quad \mathbb{E}[\tilde{\mathbf{X}}|\mathbf{Y} = \mathbf{y}] = \mathbf{0}. \quad (\text{A.8})$$

*Important:* if  $\hat{x}_i^{\text{S}}$  is the symbolwise estimate of  $x_i$  with the smallest expected square error, then  $(\hat{x}_1^{\text{S}}, \dots, \hat{x}_n^{\text{S}})^T = \hat{\mathbf{x}}^{\text{S}}$  (where  $\hat{\mathbf{x}}^{\text{S}}$  is the blockwise estimate).

## A.2 A Second Family of Estimators

### A.2.1 Blockwise MAP Estimator

The blockwise maximum a-posteriori (MAP) estimate

$$\hat{\mathbf{x}}^{\text{MAP}} = \mathbf{g}^{\text{MAP}}(\mathbf{y}) \quad (\text{A.9})$$

is the value of  $\mathbf{x}$  that maximizes the conditional density

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}), \quad (\text{A.10})$$

or, equivalently, that maximizes

$$p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}). \quad (\text{A.11})$$

### A.2.2 Symbolwise MAP Estimator

For each  $i = 1, \dots, n$ , the symbolwise maximum a-posteriori (MAP) estimate

$$\hat{x}_i^{\text{MAP}} = \mathbf{g}_i^{\text{MAP}}(\mathbf{y}) \quad (\text{A.12})$$

is the value of  $x_i$  that maximizes the conditional density

$$p_{X_i|\mathbf{Y}}(x_i|\mathbf{y}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_{\mathbf{X}|\mathbf{Y}}(x'_1, \dots, x'_{i-1}, x_i, x'_{i+1}, \dots, x'_n|\mathbf{y}) dx'_1 \cdots dx'_{i-1} dx'_{i+1} \cdots dx'_n. \quad (\text{A.13})$$

This is equivalent to taking that  $x_i$  which maximizes

$$p_{X_i\mathbf{Y}}(x_i, \mathbf{y}). \quad (\text{A.14})$$

### A.2.3 Uniform Cost Estimator

Let the uniform cost function be defined as

$$C^{U,\varepsilon}(\tilde{\mathbf{x}}) := \begin{cases} 0 & (\text{if } |\tilde{x}_i| < \varepsilon/2 \text{ for } i = 1, \dots, n) \\ 1 & (\text{otherwise}) \end{cases} \quad (\text{A.15})$$

Then

$$\hat{\mathbf{x}}^{U,\varepsilon} = \mathbf{g}^{U,\varepsilon}(\mathbf{y}), \quad (\text{A.16})$$

where  $\mathbf{g}^{U,\varepsilon}$  is the estimator  $\mathbf{g}$  that minimizes

$$\mathbb{E}[C^{U,\varepsilon}(\mathbf{X} - \mathbf{g}(\mathbf{Y}))]. \quad (\text{A.17})$$

### A.2.4 Uniform Cost Function (in the Limit)

This is the uniform cost estimator in the limit as  $\varepsilon$  tends to zero

$$\hat{\mathbf{x}}^U = \mathbf{g}^{U,\varepsilon}(\mathbf{y}) := \lim_{\varepsilon \rightarrow 0} \mathbf{g}^{U,\varepsilon}(\mathbf{y}). \quad (\text{A.18})$$

### A.2.5 Properties of the Second Family of Estimators

Practically, it holds that

$$\mathbf{g}^U = \mathbf{g}^{\text{MAP}} \quad (\text{A.19})$$

*Important:* if  $\hat{x}_i^{\text{MAP}}$  is the symbolwise MAP estimate of  $x_i$  and  $\hat{\mathbf{x}}^{\text{MAP}}$  is the blockwise MAP estimate of  $\mathbf{x}$ , then  $(\hat{x}_1^{\text{MAP}}, \dots, \hat{x}_n^{\text{MAP}})^T \neq \hat{\mathbf{x}}^{\text{MAP}}$  in general.

## A.3 Cramér-Rao Bound

Define the the so-called Fisher information matrix  $n \times n$ -matrix  $\mathbf{J}$  with components

$$J_{i,j} := \mathbb{E} \left[ \left( \frac{\partial}{\partial x_i} \ln p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \right) \left( \frac{\partial}{\partial x_j} \ln p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \right) \right] = \mathbb{E} \left[ -\frac{\partial^2}{\partial x_i \partial x_j} \ln p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \right] \quad (\text{A.20})$$

For estimates for which

$$\lim_{\mathbf{x} \rightarrow \pm\infty} \mathbb{E}[\hat{\mathbf{X}}(\mathbf{Y}) - \mathbf{X} | \mathbf{X} = \mathbf{x}] \cdot p_{\mathbf{X}}(\mathbf{x}) = \mathbf{0} \quad (\text{A.21})$$

holds, the Cramér-Rao bound is

$$\mathbb{E}[(\hat{\mathbf{X}} - \mathbf{X})(\hat{\mathbf{X}} - \mathbf{X})^T] \geq \mathbf{J}^{-1}, \quad (\text{A.22})$$

with equality for jointly Gaussian random variables. ( $\mathbf{A} \geq \mathbf{B}$  means that  $\mathbf{A} - \mathbf{B}$  is a non-negative definite matrix.)

## A.4 Jointly Gaussian Random Variables

Assume that  $\mathbf{X}$  and  $\mathbf{Y}$  are jointly Gaussian. Then

$$\mathbf{g}^E = \mathbf{g}^{S, \mathbf{Q}} = \mathbf{g}^S = \mathbf{g}^{\text{BFMV}} = \mathbf{g}^U = \mathbf{g}^{\text{MAP}} \quad (\text{independent of } \mathbf{Q}). \quad (\text{A.23})$$

*Important:* In the jointly Gaussian case, not only  $(\hat{x}_1^S, \dots, \hat{x}_n^S)^T = \hat{\mathbf{x}}^S$ , but also  $(\hat{x}_1^{\text{MAP}}, \dots, \hat{x}_n^{\text{MAP}})^T = \hat{\mathbf{x}}^{\text{MAP}}$  holds.

As mentioned in Section A.3, the Cramér-Rao bound for jointly Gaussian random variables is exact (i.e. an equality).



## Appendix B

# The Node-Potentials Method

### B.1 From a Linear Electrical Network to a Linear System of Equations

We start with the electrical network in Fig. B.1 where the  $G'_{i,j}$ 's are conductances, the  $U_i$ 's are voltages and  $I_{q,i}$ 's are current sources. To simplify notation afterwards, we define  $G'_{j,i} := G'_{i,j}$  when  $j > i$ . Writing the equations corresponding to the Kirchhoff current law, we obtain

$$G'_{1,1}(U_1 - 0) + G'_{1,2}(U_1 - U_2) + G'_{1,3}(U_1 - U_3) = I_{q,1}, \quad (\text{B.1})$$

$$G'_{2,1}(U_2 - U_1) + G'_{2,2}(U_2 - 0) + G'_{2,3}(U_2 - U_3) = I_{q,2}, \quad (\text{B.2})$$

$$G'_{3,1}(U_3 - U_1) + G'_{3,2}(U_3 - U_2) + G'_{3,3}(U_3 - 0) = I_{q,3}. \quad (\text{B.3})$$

This is equivalent to the system of equation

$$\mathbf{G} \cdot \mathbf{U} = \mathbf{I}_q, \quad (\text{B.4})$$

or explicitly,

$$\begin{pmatrix} G_{1,1} & G_{1,2} & G_{1,3} \\ G_{2,1} & G_{2,2} & G_{2,3} \\ G_{3,1} & G_{3,2} & G_{3,3} \end{pmatrix} \cdot \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} I_{q,1} \\ I_{q,2} \\ I_{q,3} \end{pmatrix}, \quad (\text{B.5})$$

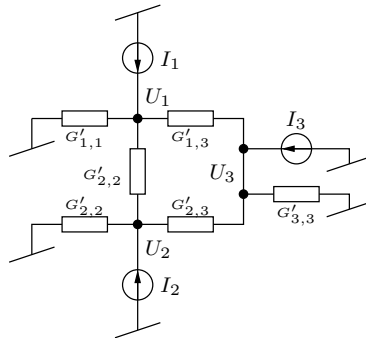


Figure B.1: Example 1.

where we have introduced

$$\begin{pmatrix} G_{1,1} & G_{1,2} & G_{1,3} \\ G_{2,1} & G_{2,2} & G_{2,2} \\ G_{3,1} & G_{3,2} & G_{3,3} \end{pmatrix} := \begin{pmatrix} G'_{1,1} + G'_{1,2} + G'_{1,3} & -G'_{1,2} & -G'_{1,3} \\ -G'_{2,1} & G'_{2,1} + G'_{2,2} + G'_{2,3} & -G'_{2,3} \\ -G'_{3,1} & -G'_{3,2} & G'_{3,1} + G'_{3,2} + G'_{3,3} \end{pmatrix}. \quad (\text{B.6})$$

Obviously,  $\mathbf{G}$  is symmetric, i.e.,  $\mathbf{G}^T = \mathbf{G}$ .

## B.2 From a Linear System of Equations to a Linear Electrical Network

Inversely, starting from a system of equations as in Eq. (B.4), or equivalently as in Eq. (B.5), we can derive an electrical network as in Fig. B.1 where the conductances are

$$\begin{pmatrix} G'_{1,1} & G'_{1,2} & G'_{1,3} \\ G'_{2,1} & G'_{2,2} & G'_{2,2} \\ G'_{3,1} & G'_{3,2} & G'_{3,3} \end{pmatrix} := \begin{pmatrix} G_{1,1} + G_{1,2} + G_{1,3} & -G_{1,2} & -G_{1,3} \\ -G_{2,1} & G_{2,1} + G_{2,2} + G_{2,3} & -G_{2,3} \\ -G_{3,1} & -G_{3,2} & G_{3,1} + G_{3,2} + G_{3,3} \end{pmatrix}. \quad (\text{B.7})$$

The general case with  $n$  equations follows easily from the above example where  $n = 3$ . The resulting electrical network has  $n$  nodes plus a ground node. More information about this method and its dual, the loop-currents method, can be found e.g. [17].

## Appendix C

# Lagrange Duality

The aim of this appendix is to give a short introduction into Lagrange duality. Some of the topics can also be found e.g. in [1] and [19]. Instead of using  $\mathbf{y}$  and  $\mathbf{z}$  in the context of conjugate functions as in [1], we used  $\boldsymbol{\xi}$  and  $\boldsymbol{\gamma}$ , respectively, to avoid overlapping notation with the other chapters of this report. Throughout this chapter, by convexity we mean strict convexity; basically, all results can also be carried over to functions that are convex (but not strictly convex), although one has to be more careful.<sup>1</sup> We follow this simplified path, because this is enough for getting the main idea of Lagrange duality.

Of course, every statement about a convex function can be – with the appropriate changes – turned into a statement about a concave function.

We give a list of the variables and functions that will appear in this appendix and in Sec. 6 and state the relations among them. By a generalized distribution we mean a function that properly scaled could be a distribution.

$\mathbf{x}$	variables	$\boldsymbol{\xi}$	dual variables
$f(\cdot)$	generalized distribution	$f^*(\cdot)$	dual generalized distribution
$\phi(\cdot)$	objective function	$\phi^*(\cdot)$	conjugate function
$s(\cdot)$	score function	$s^*(\cdot)$	dual score function

The relation among the different functions is as follows. (The most important is the first one, the others are defined relative to  $\phi(\cdot)$  and  $\phi^*(\cdot)$ .)

- $\phi^*(\cdot)$  is the conjugate function of  $\phi(\cdot)$  and vice-versa,
- $f(\cdot) := \exp(-\phi(\cdot))$  and  $\phi(\cdot) = -\ln f(\cdot)$ ,
- $f^*(\cdot) := \exp(-\phi^*(\cdot))$  and  $\phi^*(\cdot) = -\ln f^*(\cdot)$ ,
- $s(\mathbf{x}) := \frac{\partial}{\partial \mathbf{x}} \phi(\mathbf{x})$ ,
- $s^*(\boldsymbol{\xi}) := \frac{\partial}{\partial \boldsymbol{\xi}} \phi^*(\boldsymbol{\xi})$ .

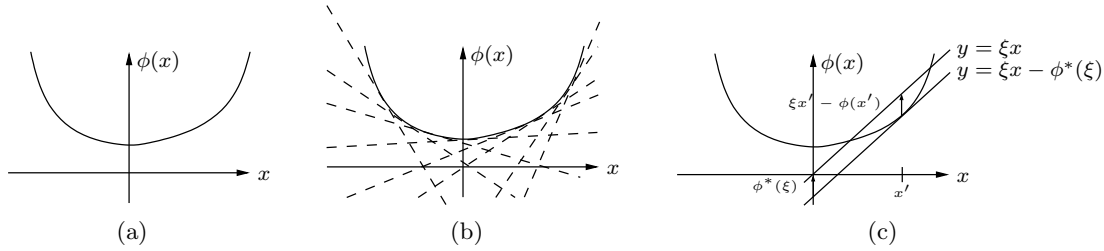


Figure C.1: Left: convex function  $\phi(\cdot)$ . Middle: the function  $\phi(\cdot)$  as envelope. Right: deriving the function  $\phi^*(\cdot)$ . (The arrows indicate in which direction the differences are counted positively.)

## C.1 Convex Functions and Their Conjugate Function

This section introduces the conjugate function<sup>2</sup> of a convex function. First, the one-dimensional case is stated and proven, from which the generalization to higher dimensions easily follows. The main observation is that tangents to convex functions are always below or on the curve, but never above. The idea is to describe the curve by a family of tangents, i.e. the curve shall be the envelope of a tangent family. The family parameter will be the slope, and the negative conjugate function will say for each slope where the corresponding tangent crosses the vertical axis.

**Theorem 4** Let  $\phi : \mathcal{D}_\phi \rightarrow \mathbb{R}$  be a convex function ( $\mathcal{D}_\phi \subseteq \mathbb{R}$ ). This function can be written as

$$\phi(x) = \max_{\xi \in \mathcal{D}_{\phi^*}} (\xi x - \phi^*(\xi)), \quad (\text{C.1})$$

where  $\phi^* : \mathcal{D}_{\phi^*} \rightarrow \mathbb{R}$  is the conjugate function ( $\mathcal{D}_{\phi^*} \subseteq \mathbb{R}$ )

$$\phi^*(\xi) = \max_{x \in \mathcal{D}_\phi} (\xi x - \phi(x)), \quad (\text{C.2})$$

*Proof:* Fig. C.1(a) depicts a convex function  $y = \phi(x)$ , and Fig. C.1(b) shows that this function is the envelope of the family of lines  $y = \xi x - \phi^*(\xi)$ , where the slope  $\xi$  is the family parameter. Fig. C.1(c) hints at the main idea of the proof.

Assume that a family of linear functions is defined in such a manner that the envelope is  $\phi(x) = \max_{\xi \in \mathcal{D}_{\phi^*}} (\xi x - \phi^*(\xi))$ . Recognizing the parallelogram in Fig. C.1(c), the conjugate function  $\phi^*(a)$  can readily be seen to be  $\phi^*(\xi) = \max_{x \in \mathcal{D}_\phi} (\xi x - \phi(x))$ .  $\square$

**Theorem 5** Let  $\phi : \mathcal{D}_\phi \rightarrow \mathbb{R}$  be a convex function ( $\mathcal{D}_\phi \subseteq \mathbb{R}^n$ ,  $n \geq 1$ ). This function can be written as

$$\phi(\mathbf{x}) = \max_{\boldsymbol{\xi} \in \mathcal{D}_{\phi^*}} (\boldsymbol{\xi}^T \mathbf{x} - \phi^*(\boldsymbol{\xi})), \quad (\text{C.3})$$

<sup>1</sup>For easiness of exposition, we also assume the functions to be differentiable; the results can be extended to continuous (but not differentiable) functions.

<sup>2</sup>What we call the conjugate function is often also called the Legendre transform (sometimes with a minus in front).

where  $\phi^* : \mathcal{D}_{\phi^*} \rightarrow \mathbb{R}$  is the conjugate function ( $\mathcal{D}_{\phi^*} \subseteq \mathbb{R}^n$ )

$$\phi^*(\boldsymbol{\xi}) = \max_{\mathbf{x} \in \mathcal{D}_{\phi}} (\boldsymbol{\xi}^T \mathbf{x} - \phi(\mathbf{x})), \quad (\text{C.4})$$

*Proof:* The idea is similar to the one in the proof of Theorem 4. The slope is generalized to the gradient of the function  $\phi(\cdot)$ .  $\square$

**Theorem 6** *The conjugate function  $\phi^*(\boldsymbol{\xi})$  of a convex function  $\phi(\mathbf{x})$  is convex. The conjugate of the conjugate function is the original function.*

*Proof:* We prove it here for the one-dimensional case. To determine the value of  $\phi^*(\xi)$  for a certain  $\xi$  one sets the derivative of  $\xi x - \phi(x)$  with respect to  $x$  equal to zero.

$$\frac{\partial}{\partial x}(\xi x - \phi(x)) = \xi - \frac{\partial}{\partial x}\phi(x) = \xi - s(x) \stackrel{!}{=} 0, \quad (\text{C.5})$$

where  $s(x) := \frac{\partial}{\partial x}\phi(x)$  is the score function. As  $\phi(\cdot)$  is convex,  $s(\cdot)$  is monotonically increasing. On the other hand, to determine the value of  $\phi(x)$  for a certain  $x$  one derives  $\xi x - \phi^*(\xi)$  with respect to  $\xi$  and sets this equal to zero.

$$\frac{\partial}{\partial \xi}(\xi x - \phi^*(\xi)) = x - \frac{\partial}{\partial \xi}\phi^*(\xi) = x - s^*(\xi) \stackrel{!}{=} 0, \quad (\text{C.6})$$

where  $s^*(\xi) := \frac{\partial}{\partial \xi}\phi^*(\xi)$  is the dual score function. Combining Eqs. (C.5) and (C.6) we see that

$$s(x) = (s^*)^{-1}(x), \quad s^*(\xi) = s^{-1}(\xi), \quad (\text{C.7})$$

i.e.  $s(\cdot)$  and  $s^*(\cdot)$  are a pair of inverse functions. Because  $s(\cdot)$  is monotonically increasing,  $s^*(\cdot)$  must also be monotonically increasing which implies the convexity of  $\phi^*(\cdot)$ .

Alternatively, contemplating Fig. C.1(b), one can see that this holds for the one-dimensional case.

From this proof it follows also directly that the conjugate of the conjugate function is the original function again.  $\square$

### C.1.1 An Example

We give a little example of calculating the conjugate function (Legendre transformation). We consider the quadratic function

$$\phi(x) = \frac{1}{2\beta}(x - x_0)^2. \quad (\text{C.8})$$

The score function is

$$s(x) = \frac{1}{\beta}(x - x_0), \quad (\text{C.9})$$

i.e. a function with slope  $1/\beta$  that crosses the  $x$ -axis at  $x_0$ . The conjugate function and the dual score function are, respectively,

$$\phi^*(\xi) = x\xi - \phi(x)|_{x=\beta\xi+x_0} = \frac{1}{2}\beta\xi^2 + x_0\xi = \frac{1}{2}\beta\left(\xi + \frac{x_0}{\beta}\right)^2 - \frac{x_0^2}{2\beta} \quad (\text{C.10})$$

$$s^*(\xi) = \beta\xi + x_0. \quad (\text{C.11})$$

In the limit  $\beta \rightarrow 0$  (important for considerations as in Sec. 1.6) the score function  $s(\cdot)$  is a vertical function that crosses the  $x$ -axis at  $x_0$ , the Lagrange dual is  $\phi^*(\xi) = x_0\xi$  and the dual score function is  $s^*(\xi) = x_0$ .

## C.2 Duality in Equality Constrained Minimization

### C.2.1 The General Convex Minimization Problem

Let  $\phi(\mathbf{x})$  be a convex function (called the objective function) which we would like to minimize under the constraint  $\mathbf{Ax} = \mathbf{b}$ , where  $\mathbf{x}$  is  $n$ -dimensional,  $\mathbf{A}$  is an  $m \times n$ -matrix and  $\mathbf{b}$  is an  $m \times 1$ -vector, i.e. we are interested in

$$\min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}). \quad (\text{C.12})$$

The Lagrangian of this problem is  $L = \phi(\mathbf{x}) - \boldsymbol{\lambda}^T(\mathbf{Ax} - \mathbf{b})$  and deriving it with respect to  $\mathbf{x}$  and  $\boldsymbol{\lambda}$ , respectively, and setting equal to zero, we obtain

$$\frac{\partial}{\partial \mathbf{x}} \phi(\mathbf{x}) - \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0}, \quad (\text{C.13})$$

$$-\mathbf{Ax} + \mathbf{b} = \mathbf{0}. \quad (\text{C.14})$$

To reformulate Eq. (C.12), we express  $\phi(\cdot)$  with the help of the conjugate function  $\phi^*(\boldsymbol{\xi})$  as shown in Theorem 5

$$\min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax}=\mathbf{b}} \max_{\boldsymbol{\xi} \in \mathbb{R}^n} (\boldsymbol{\xi}^T \mathbf{x} - \phi^*(\boldsymbol{\xi})). \quad (\text{C.15})$$

But in Eq. (C.13) we have seen that the ‘‘slope’’  $\partial\phi(\mathbf{x})/\partial\mathbf{x}$  (the gradient of  $\phi(\cdot)$ ) at the solution point must be  $\mathbf{A}^T \boldsymbol{\lambda}$ , i.e. in the column-space of  $\mathbf{A}^T$ , or equivalently in the row-space of  $\mathbf{A}$ . Therefore, we do not change the solution of Eq. (C.15) by restricting the slope (the gradient)  $\boldsymbol{\xi}$  to lie in the column-space of  $\mathbf{A}^T$ :  $\boldsymbol{\xi} = \mathbf{A}^T \boldsymbol{\gamma}$  for some  $m \times 1$ -vector  $\boldsymbol{\gamma}$ . We conclude that

$$\min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax}=\mathbf{b}} \phi(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax}=\mathbf{b}} \max_{\boldsymbol{\gamma} \in \mathbb{R}^m} (\boldsymbol{\gamma}^T \mathbf{Ax} - \phi^*(\mathbf{A}^T \boldsymbol{\gamma})) \quad (\text{C.16})$$

$$= \min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{Ax}=\mathbf{b}} \max_{\boldsymbol{\gamma} \in \mathbb{R}^m} (\boldsymbol{\gamma}^T \mathbf{b} - \phi^*(\mathbf{A}^T \boldsymbol{\gamma})) \quad (\text{C.17})$$

$$= \max_{\boldsymbol{\gamma} \in \mathbb{R}^m} (\boldsymbol{\gamma}^T \mathbf{b} - \phi^*(\mathbf{A}^T \boldsymbol{\gamma})) \quad (\text{C.18})$$

$$= \max_{\boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\gamma} \in \mathbb{R}^m: \mathbf{A}^T \boldsymbol{\gamma}=\boldsymbol{\xi}} (\boldsymbol{\gamma}^T \mathbf{b} - \phi^*(\boldsymbol{\xi})). \quad (\text{C.19})$$

This proves the next theorem (which could also have been derived from Lagrange duality or Fenchel duality<sup>3</sup>).

<sup>3</sup>As has been shown in [20], Fenchel duality and Lagrange duality are equivalent.

**Theorem 7** Let  $\phi(\cdot)$  be an  $n$ -dimensional convex function and let  $\phi^*(\cdot)$  be its conjugate function (which is also an  $n$ -dimensional convex function). Then we have equality of the minimization and maximization problems

$$\min_{x \in \mathbb{R}^n: \mathbf{A}\mathbf{x}=\mathbf{b}} \phi(\mathbf{x}) = \max_{\xi \in \mathbb{R}^n, \gamma \in \mathbb{R}^m: \mathbf{A}^T\gamma=\xi} (\gamma^T\mathbf{b} - \phi^*(\xi)) = \max_{\gamma \in \mathbb{R}^m} (\gamma^T\mathbf{b} - \phi^*(\mathbf{A}^T\gamma)). \quad (\text{C.20})$$

### C.2.2 The Quadratic Minimization Problem

(This section is based on [1].) Instead of the general convex function  $\phi(\cdot)$  in Sec. C.2.1, let us now consider the specific example  $\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x}$ , where we assume that  $\mathbf{P}$  is positive definite, so its inverse  $\mathbf{P}^{-1}$  exists. Then solving the minimization problem stated as the “Primal Quadratic Minimum Problem” is equivalent to solving the “Quadratic Lagrangian Problem”.

- **Primal Quadratic Minimum Problem**

Minimize (over  $\mathbf{x}$ )

$$\frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} \quad (\text{C.21})$$

with

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \quad (\text{C.22})$$

- **Quadratic Lagrangian Problem**

Solve

$$\mathbf{A}^T\gamma - \mathbf{P}\mathbf{x} = \mathbf{c} \quad (\text{C.23})$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \quad (\text{C.24})$$

- **Dual Quadratic Maximum Problem**

Maximize (over  $\xi$  and  $\gamma$ )

$$-\frac{1}{2}\xi^T\mathbf{P}^{-1}\xi + \mathbf{b}^T\gamma \quad (\text{C.25})$$

with

$$\mathbf{A}^T\gamma - \xi = \mathbf{c}. \quad (\text{C.26})$$

- **Quadratic Lagrangian Problem**

Solve

$$\mathbf{A}^T\gamma - \xi = \mathbf{c} \quad (\text{C.27})$$

$$\mathbf{x} = \mathbf{P}^{-1}\xi \quad (\text{C.28})$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \quad (\text{C.29})$$

*Proof:* (From the “Primal Quadratic Minimum Problem” to its “Quadratic Lagrangian Problem”) Let  $\gamma$  be the vector with the Lagrange multipliers. The Lagrangian function is  $L = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} - \gamma^T(\mathbf{A}\mathbf{x} - \mathbf{b})$ . Deriving  $L$  with respect to  $\mathbf{x}$  and setting equal to zero yields  $\mathbf{P}\mathbf{x} + \mathbf{c} - \mathbf{A}^T\gamma \stackrel{!}{=} \mathbf{0}$  or  $\mathbf{A}^T\gamma - \mathbf{P}\mathbf{x} = \mathbf{c}$ . Deriving  $L$  with respect to  $\gamma$  gives the constraint  $\mathbf{A}\mathbf{x} = \mathbf{b}$  and we get the quadratic Lagrangian problem stated above.  $\square$

*Proof:* (From the “Dual Quadratic Maximum Problem” to its “Quadratic Lagrangian Problem”) Let  $\mathbf{x}$  be the vector with the Lagrange multipliers. The Lagrangian function is  $L = -\frac{1}{2}\boldsymbol{\xi}^T\mathbf{P}^{-1}\boldsymbol{\xi} + \mathbf{b}^T\boldsymbol{\gamma} - \mathbf{x}^T(\mathbf{A}^T\boldsymbol{\gamma} - \boldsymbol{\xi} - \mathbf{c})$ . Derivation of  $L$  firstly with respect to  $\boldsymbol{\xi}$  and setting equal to zero yields  $-\mathbf{P}^{-1}\boldsymbol{\xi} + \mathbf{x} \stackrel{!}{=} \mathbf{0}$  or  $\mathbf{x} = \mathbf{P}^{-1}\boldsymbol{\xi}$ . Derivation of  $L$  secondly with respect to  $\boldsymbol{\gamma}$  and setting equal to zero yields  $\mathbf{b} - \mathbf{A}\mathbf{x} \stackrel{!}{=} \mathbf{0}$  or  $\mathbf{A}\mathbf{x} = \mathbf{b}$ . Finally, deriving  $L$  with respect to  $\mathbf{x}$  and setting equal to zero, we obtain the equality constraint  $\mathbf{A}^T\boldsymbol{\gamma} - \boldsymbol{\xi} = \mathbf{c}$ ; altogether, we get the quadratic Lagrangian problem.  $\square$

The maximization problem was stated so that it yields the same quadratic Lagrangian problem as for the primal problem. This is of course nothing else than an application of Theorem 7. Indeed, the dual function of

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} \quad (\text{C.30})$$

is

$$\phi^*(\boldsymbol{\xi}) = \max_{\mathbf{x} \in \mathbb{R}^n} (\boldsymbol{\xi}^T\mathbf{x} - \phi(\mathbf{x})) = \boldsymbol{\xi}^T\mathbf{x} - \phi(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{P}^{-1}(\boldsymbol{\xi}-\mathbf{c})} \quad (\text{C.31})$$

$$= \frac{1}{2}(\boldsymbol{\xi} - \mathbf{c})^T\mathbf{P}^{-1}(\boldsymbol{\xi} - \mathbf{c}). \quad (\text{C.32})$$

Applying Theorem 7 we get

$$\min_{\mathbf{x} \in \mathbb{R}^n: \mathbf{A}\mathbf{x}=\mathbf{b}} \phi(\mathbf{x}) = \max_{\boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\gamma} \in \mathbb{R}^m: \mathbf{A}^T\boldsymbol{\gamma}=\boldsymbol{\xi}} (\boldsymbol{\gamma}^T\mathbf{b} - \phi^*(\boldsymbol{\xi})) \quad (\text{C.33})$$

$$= \max_{\boldsymbol{\xi}' \in \mathbb{R}^n, \boldsymbol{\gamma} \in \mathbb{R}^m: \mathbf{A}^T\boldsymbol{\gamma}=\boldsymbol{\xi}'+\mathbf{c}} (\boldsymbol{\gamma}^T\mathbf{b} - \phi^*(\boldsymbol{\xi}'+\mathbf{c})) \quad (\text{C.34})$$

$$= \max_{\boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\gamma} \in \mathbb{R}^m: \mathbf{A}^T\boldsymbol{\gamma}=\boldsymbol{\xi}+\mathbf{c}} \left( \boldsymbol{\gamma}^T\mathbf{b} - \frac{1}{2}\boldsymbol{\xi}^T\mathbf{P}^{-1}\boldsymbol{\xi} \right), \quad (\text{C.35})$$

which is the “Dual Quadratic Maximization Problem” in Eqs. (C.25) and (C.26).

### C.3 Electrical Networks with Voltage Sources, Current Sources and Resistors

The book by Dennis [1] shows implementations of electrical networks which solve the quadratic minimal problem of Sec. C.2.2 using ideal voltage sources, ideal current sources, resistors, and DC-transformers. Dennis also considers electrical networks solving the dual problem. In both cases one gets the same electrical network but in the first case one optimizes over the currents whereas in the second case over the voltages. Whereas the Lagrange multipliers turn out to be node potentials in the first case, they are branch currents in the second case.



## C.4 Max-Product Algorithm and Lagrange Duality

In this section we would like to find out what Lagrange duality means in the context of the MPA (max-product algorithm).<sup>4</sup>

### C.4.1 The Summation-Function Node

The MPA update rule for a summation function node as in Fig. 4.6(left), assuming  $x_3 = x_1 + x_2$  is

$$\mu_{f \rightarrow X_3}(x_3) \propto \max_{x_1, x_2} f(x_1, x_2, x_3) \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(x_2) \quad (\text{C.36})$$

$$= \max_{x_1, x_2} \delta(x_3 - x_1 - x_2) \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(x_2) \quad (\text{C.37})$$

$$\propto \max_{x_1, x_2} [x_3 - x_1 - x_2 = 0] \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(x_2) \quad (\text{C.38})$$

$$= \max_{x_1 + x_2 = x_3} \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(x_2) \quad (\text{C.39})$$

$$= \exp \left( \max_{x_1, x_2: x_1 + x_2 = x_3} (\ln \mu_{X_1 \rightarrow f}(x_1) + \ln \mu_{X_2 \rightarrow f}(x_2)) \right) \quad (\text{C.40})$$

$$= \exp \left( - \min_{x_1, x_2: x_1 + x_2 = x_3} (-\ln \mu_{X_1 \rightarrow f}(x_1) - \ln \mu_{X_2 \rightarrow f}(x_2)) \right) \quad (\text{C.41})$$

$$= \exp \left( - \min_{\mathbf{x}: \mathbf{A}\mathbf{x}=\mathbf{b}} g(\mathbf{x}) \right), \quad (\text{C.42})$$

with

$$\phi(\mathbf{x}) = \phi_1(x_1) + \phi_2(x_2), \quad (\text{C.43})$$

$$\phi(x_1) = -\ln \mu_{X_1 \rightarrow f}(x_1), \quad \phi(x_2) = -\ln \mu_{X_2 \rightarrow f}(x_2) \quad (\text{C.44})$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{b} = (x_3). \quad (\text{C.45})$$

The variable names have been chosen so as to highlight their connection to the minimization problem in Th. 7. Reformulating the minimization in Eq. (C.42) using the dual problem according to Th. 7 we obtain

$$\mu_{f \rightarrow X_3}(x_3) = \exp \left( - \max_{\boldsymbol{\xi}, \boldsymbol{\gamma}: \mathbf{A}^T \boldsymbol{\gamma} = \boldsymbol{\xi}} (\boldsymbol{\gamma}^T \mathbf{b} - \phi^*(\boldsymbol{\xi})) \right) \quad (\text{C.46})$$

$$= \exp \left( - \max_{\boldsymbol{\gamma}} (\boldsymbol{\gamma}^T \mathbf{b} - \phi^*(\mathbf{A}^T \boldsymbol{\gamma})) \right), \quad (\text{C.47})$$

with

$$\phi^*(\boldsymbol{\xi}) = \phi_1^*(\xi_1) + \phi_2^*(\xi_2) \quad (\text{C.48})$$

and  $\boldsymbol{\xi} = (\xi_1, \xi_2)^T$  and  $\boldsymbol{\gamma} = (\gamma_1)$ . Choosing the formulation of Eq. (C.46), deriving the dual Lagrangian

$$L^* := \gamma_1 x_3 - \phi_1^*(\xi_1) - \phi_2^*(\xi_2) - \lambda_1^*(\gamma_1 - \xi_1) - \lambda_2^*(\gamma_1 - \xi_2) \quad (\text{C.49})$$

---

<sup>4</sup>In this section we follow a similar “local” approach as in App. D.

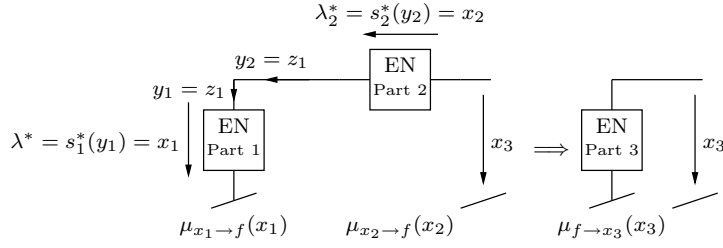


Figure C.2: Electrical network of the dual system of equations.

with respect to  $\xi$ ,  $\gamma$ , and  $\lambda^*$ , respectively, one gets

$$\left\{ \begin{array}{l} \frac{\partial}{\partial \xi_1} L^* = -\frac{\partial}{\partial \xi_1} \phi_1^*(\xi_1) + \lambda_1^* \stackrel{!}{=} 0 \quad (\text{component law}) \\ \frac{\partial}{\partial \xi_2} L^* = -\frac{\partial}{\partial \xi_2} \phi_2^*(\xi_2) + \lambda_2^* \stackrel{!}{=} 0 \quad (\text{component law}) \\ \frac{\partial}{\partial \gamma_1} L^* = x_3 - \lambda_1^* + \lambda_2^* \stackrel{!}{=} 0 \quad (\text{Kirchhoff voltage law}) \\ \frac{\partial}{\partial \lambda_1^*} L^* = -\gamma_1 + \xi_1 \stackrel{!}{=} 0 \quad (\text{Kirchhoff voltage law}) \\ \frac{\partial}{\partial \lambda_2^*} L^* = -\gamma_1 + \xi_2 \stackrel{!}{=} 0 \quad (\text{Kirchhoff voltage law}) \end{array} \right. \quad (\text{C.50})$$

This system of equations can be interpreted as the electrical network shown in Fig. C.2. Let  $s_1(x_1) = \frac{\partial}{\partial x_1} \phi_1(x_1)$  be the score function and let  $s_1^*(\xi_1) = \frac{\partial}{\partial \xi_1} \phi_1^*(\xi_1)$  be the dual score function. In the proof of Th. 6 we have seen that

$$s_1(x_1) = (s_1^*)^{-1}(x_1), \quad s_1^*(\xi_1) = s_1^{-1}(\xi_1), \quad (\text{C.51})$$

i.e.  $s_1(\cdot)$  and  $s_1^*(\cdot)$  are a pair of inverse functions. In electrical network language  $s_1(\cdot)$  is the current-voltage characteristic  $I_1 = I_1(U_1)$ , whereas  $s_1^*(\cdot)$  is the corresponding voltage-current characteristic  $U_1 = U_1(I_1)$ .

Note that in the context of electrical networks as in Ch. 3, the function  $\phi_1(\cdot)$  has an argument of dimension Volt and a result of dimension Volt<sup>2</sup>/Ohm = Power, so the the dimension of the slope of the tangents are Volt<sup>2</sup>/Ohm/Volt = Volt/Ohm = Ampère, the dimension of the dual variables. Analogously,  $\phi_1^*(\cdot)$  has an argument of dimension Ampère and a result of dimension Ohm · Ampère<sup>2</sup> = Power, so the dimension of the slope of the tangents are Ohm · Ampère<sup>2</sup>/Ampère = Ohm · Ampère = Volt, the dimension of the original variables. So the primal variables are voltages and the dual variables are currents. (Dennis [1] has chosen the primal variables to be currents and the dual variables to be voltages.)

To conclude, when using the techniques of Ch. 3, the electrical network derived from the dual problem is the same circuit as from the primal problem, but now the components are described in a in terms of voltage-current characteristics instead of current-voltage characteristics.

## C.4.2 The Equality-Function Node and the Scaling-Funtion Node

A similar approach as in Sec. C.4.1 can be taken to analyze an equality-function and a scaling-function node. Again, the topology remains the same but the parts are interpreted in terms of voltage-current characteristics and not in terms of current-voltage characteristics.

## C.5 Lagrange Dual of Subspace Constraints

We consider general constraint-function nodes with  $n$  edges connected to it. In this section we derive the dual generalized distribution of such a function for the case that the set of all valid configurations corresponds to a  $d$ -dimensional subspace. Essentially, it is again a constraint function node and the set of all its valid configurations corresponds to the orthogonal complement of the above subspace. To simplify matters we use the same scaled Gaussian distributions as at the beginning of Sec. 6.1.1.

We give the following examples for  $n = 3$ , which are obviously dual to each other.

- The equality-function node is a constraint-function node related to the 1-dimensional subspace spanned by  $(1, 1, 1)$  with orthogonal complement spanned by  $(1, 0, -1)$  and  $(0, 1, -1)$ . The Lagrange dual is therefore the summation-function node.
- The summation-function node (all arrows are pointing inwards) is a constraint function related to the 2-dimensional subspace spanned by  $(1, 0, -1)$  and  $(0, 1, -1)$  with orthogonal complement spanned by  $(1, 1, 1)$ . The Lagrange dual is therefore the equality-function node.

**Lemma 8** *Let  $f(\mathbf{x})$  be the (generalized) distribution of a constraint-function node of degree  $n$  (of the type introduced at the beginning of Sec. 6.1.1) where the set of all valid configurations is a  $d$ -dimensional subspace of  $\mathbb{R}^n$ . Then the dual generalized distribution  $f^*(\boldsymbol{\xi}_t)$  is also a constraint function node where the set of all valid configurations is the  $n - d$ -dimensional orthogonal complement of the above  $d$ -dimensional subspace. For any valid configuration  $\mathbf{x}$  of the primal function node  $f(\cdot)$  and any valid configuration  $\boldsymbol{\xi}$  of the dual function node  $f^*(\cdot)$  follows  $\mathbf{x}^T \cdot \boldsymbol{\xi} = 0$ .*

*Proof:* The proof is straightforward, but somewhat technical. Assume that  $\mathbf{x} = (x_1, \dots, x_n)$ , i.e.  $f(\mathbf{x}) = f(x_1, \dots, x_n)$ . If  $f(\mathbf{x})$  has the form of a scaled Gaussian distributions as explained at the beginning of Sec. 6.1.1, then the corresponding  $\phi(\mathbf{x}) = -\ln f(\mathbf{x})$  can be described as follows.

Define in the subspace of all valid configurations an orthonormal basis and combine it with an orthonormal basis of the  $n - d$ -dimensional orthogonal complement. Let the columns of the  $n \times n$ -matrix  $\mathbf{T}$  be the vectors of this basis, first from the subspace, then from the orthogonal complement. It follows that  $\mathbf{T}$  must be orthogonal, i.e.,  $\mathbf{T} \cdot \mathbf{T}^T = \mathbb{1}$ . Then we have

$$\phi(\mathbf{x}) = -\ln f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{K}_{\mathbf{X}}^{-1} \mathbf{x} \quad (\text{C.52})$$

$$(\text{C.53})$$

with  $\mathbf{K}_{\mathbf{X}} = \mathbf{T} \mathbf{K} \mathbf{T}^T$  and

$$\mathbf{K} = \text{diag} \left( \underbrace{\beta, \dots, \beta}_{d \text{ times}}, \underbrace{1/\beta, \dots, 1/\beta}_{n-d \text{ times}} \right). \quad (\text{C.54})$$

The conjugate function of  $\phi(\mathbf{x})$  is

$$\phi^*(\boldsymbol{\xi}) = \mathbf{x}^T \boldsymbol{\xi} - \phi(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{K}_{\mathbf{X}} \boldsymbol{\xi}} = \frac{1}{2} \boldsymbol{\xi}^T \mathbf{K}_{\mathbf{X}} \boldsymbol{\xi} = \frac{1}{2} \boldsymbol{\xi}^T (\mathbf{K}_{\mathbf{X}}^*)^{-1} \boldsymbol{\xi} \quad (\text{C.55})$$

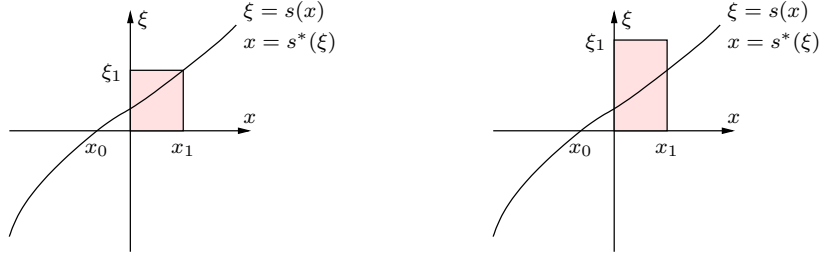


Figure C.3: Illustration to Lemma 9 about an upper bound of the shaded area.

with  $\mathbf{K}_X^* = \mathbf{TK}^*\mathbf{T}^T$  and

$$\mathbf{K}^* = \mathbf{K}^{-1} = \text{diag} \left( \underbrace{1/\beta, \dots, 1/\beta}_d \text{ times}, \underbrace{\beta, \dots, \beta}_{n-d} \text{ times} \right). \quad (\text{C.56})$$

But  $f^*(\boldsymbol{\xi}) = \exp(-\phi^*(\cdot))$  has the form of a scaled Gaussian distribution as introduced at the beginning of Sec. 6.1.1 with variance  $\beta$  in the  $n - d$ -dimensional orthogonal subspace and variance  $1/\beta$  in the  $d$ -dimensional subspace. In the limit  $\beta \rightarrow 0$  we get the desired result.  $\square$

## C.6 Fenchel's Inequality

The next lemma will first be formulated abstractly. Then we will give a more intuitive geometric proof and interpretation of its content. It is known as Fenchel's inequality (or Young's inequality if  $\phi(\cdot)$  is differentiable), see [19].

**Lemma 9** *Let  $\phi(\cdot)$  be a convex function with score function  $s(\cdot)$  and with conjugate function  $\phi^*(\cdot)$ . For any  $x_1, \xi_1$ ,*

$$x_1 \cdot \xi_1 \leq \phi(x_1) + \phi^*(\xi_1), \quad (\text{C.57})$$

where equality holds if and only if  $\xi_1 = s(x_1)$ .

*Proof:* If  $\xi_1 = s(x_1)$ , we get from the definition of the Lagrange dual  $\phi^*(\xi_1) = \max_x (x\xi_1 - \phi(x)) = x_1\xi_1 - \phi(x_1)$  that  $x_1\xi_1 = \phi(x_1) + \phi^*(\xi_1)$ .

If  $\xi_1 \neq s(x_1)$ , then  $\phi^*(\xi_1) = \max_x (x\xi_1 - \phi(x)) > x_1\xi_1 - \phi(x_1)$ , which implies  $x_1\xi_1 < \phi(x_1) + \phi^*(\xi_1)$ .  $\square$

We come now to the geometric interpretation of Lemma 9. We remind the reader that  $s(x) = \frac{\partial}{\partial x}\phi(x)$  and  $s^*(\xi) = \frac{\partial}{\partial \xi}\phi^*(\xi)$ ,  $(s^*)^{-1}(x) = s(x)$ , and  $s^{-1}(\xi) = s^*(\xi)$ . Consider the plots in Fig. C.3. If  $\xi = s(x_1)$ , i.e., the point  $(x_1, \xi_1)$  lies on the score function graph, then one has the situation as in Fig. C.3(a); if  $\xi \neq s(x_1)$ , i.e., the point  $(x_1, \xi_1)$  does not lie on the score function graph, then one has the situation as in Fig. C.3(b). We now give a geometric proof for the first situation when  $(x_1, \xi_1)$  lies in the first quadrant. We introduce the point  $(x_0, \xi_0) = (x_0, 0)$  which lies on the score function graph. On the one hand the shaded area is obviously  $x_1 \cdot \xi_1$ . On the other hand, the area between the  $x$ -axis and the curve  $s(x)$  and between  $x_0$  and  $x_1$  is  $\int_{x_0}^{x_1} s(x) dx = \phi(x_1) - \phi(x_0)$ ; the (signed) area between  $\xi$ -axis and  $s^*(\xi)$

---

and between  $\xi_0 = 0$  and  $\xi_1$  is  $\int_{\xi_0}^{\xi_1} s^*(\xi) d\xi = \phi^*(\xi_1) - \phi^*(\xi_0)$ . Adding these two integrals gives the shaded area (as the area to the left of the  $\xi$ -axis is counted once positively and once negatively)  $x_1 \cdot \xi_1 = \phi(x_1) + \phi^*(\xi_1) + c$ , with  $c = -(\phi(x_0) + \phi^*(\xi_0))$ . One can show that  $c = 0$ . For the inequality statement one can give a geometric proof along the same lines.



# Appendix D

## Fourier Duality

In this appendix we consider the SPA and not the MPA; but as stated before, in the case of jointly Gaussian random variables, SPA and MPA are equal (see e.g. [8]).

In contrast to [7] and [18] which treat the Fourier-type duality of factor graphs and Forney-style factor graphs (FFGs) in a “top-down” fashion, we will give here a short introduction to this topic in an “bottom-up” fashion.<sup>1</sup>

### D.1 Dualization of the Summation-Function Node

Consider the summation-function node  $f(\cdot)$  as given in Fig. D.1(left). Note that the constraint here is defined as  $x_1 + x_2 + x_3 = 0$ . Taking the incoming messages  $\mu_{X_1 \rightarrow f}(x_1)$  and  $\mu_{X_2 \rightarrow f}(x_2)$ , the SPA calculates the outgoing message  $\mu_{f \rightarrow X_3}(x_3)$  by

$$\mu_{f \rightarrow X_3}(x_3) := \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2, x_3) \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(x_2) dx_2 dx_1 \quad (\text{D.1})$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x_1 + x_2 + x_3) \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(x_2) dx_2 dx_1 \quad (\text{D.2})$$

$$= \int_{-\infty}^{+\infty} \mu_{X_1 \rightarrow f}(x_1) \mu_{X_2 \rightarrow f}(-x_3 - x_1) dx_1. \quad (\text{D.3})$$

<sup>1</sup>This approach is similar to the one taken in Sec. C.4.

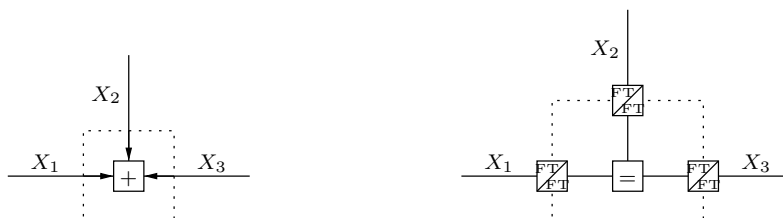


Figure D.1: Left: Summation-Function Node; Right: Calculating the messages in the Fourier domain.

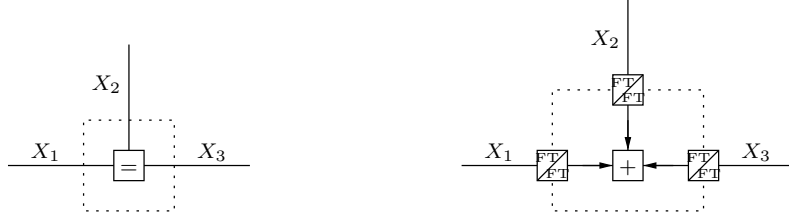


Figure D.2: Left: Equality-Function Node: Right: Calculating the messages in the Fourier domain.

This is a type of convolution. Introducing the Fourier transform  $\bar{\mu}(\bar{x})$  of a message  $\mu(x)$

$$\bar{\mu}(x) = \int_{-\infty}^{+\infty} \mu(x) \exp(-j\bar{x}x) dx \quad (\text{D.4})$$

$$\mu(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \bar{\mu}(\bar{x}) \exp(+j\bar{x}x) d\bar{x}, \quad (\text{D.5})$$

one can modify Eq. (D.3) to

$$\mu_{f \rightarrow X_3}(x_3) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \bar{\mu}_{X_1 \rightarrow f}(\bar{x}_1) \exp(+j\bar{x}_1 x_1) d\bar{x}_1 \mu_{X_2 \rightarrow f}(-x_3 - x_1) dx_1 \quad (\text{D.6})$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \bar{\mu}_{X_1 \rightarrow f}(\bar{x}_1) \int_{-\infty}^{+\infty} \mu_{X_2 \rightarrow f}(-x_3 - x_1) \exp(+j\bar{x}_1 x_1) dx_1 d\bar{x}_1 \quad (\text{D.7})$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \bar{\mu}_{X_1 \rightarrow f}(\bar{x}_1) \bar{\mu}_{X_2 \rightarrow f}(\bar{x}_1) \exp(-j\bar{x}_1 x_3) d\bar{x}_1. \quad (\text{D.8})$$

We conclude that

$$\mu_{f \rightarrow X_3}(x_3) \propto \int_{-\infty}^{+\infty} \underbrace{\bar{\mu}_{X_1 \rightarrow f}(\bar{x}) \bar{\mu}_{X_2 \rightarrow f}(\bar{x})}_{(*)} \exp(-j\bar{x}x_3) d\bar{x}. \quad (\text{D.9})$$

Such an equation could have also be derived for the outgoing messages  $\mu_{f \rightarrow X_1}$  and  $\mu_{f \rightarrow X_2}$ . Fig. D.1(right)<sup>2</sup> tries to capture these calculations: the FT/FT-box means that the messages going through must be Fourier transformed in each direction.<sup>3</sup> As can be seen from the (\*)-part of Eq. (D.9), the processing of the Fourier transformed messages corresponds to the update rules for an equality-function node.

## D.2 Dualization of the Equality-Function Node

Paralleling the whole derivation for an equality-function node, the analog to Eq. (D.9) would be

$$\mu_{f \rightarrow X_3}(x_3) \propto \int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{+\infty} \bar{\mu}_{X_1 \rightarrow f}(\bar{x}') \bar{\mu}_{X_2 \rightarrow f}(-\bar{x} - \bar{x}') d\bar{x}'}_{(*)} \exp(-j\bar{x}x_3) d\bar{x}. \quad (\text{D.10})$$

<sup>2</sup>Note that this is not anymore a proper FFG in the sense that the FT/FT nodes are not terms of a global function.

<sup>3</sup>It is indeed a Fourier transform in both direction, and not an FT/IFT pair.



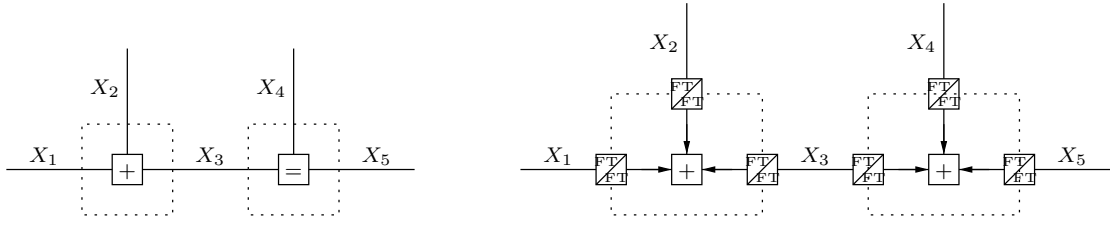


Figure D.3: Left: FFG with summation-function node and equality-function node. Right: Fourier dualization of each function node separately.

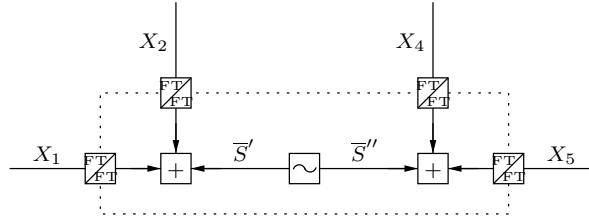


Figure D.4: Continuation of the example in Fig. D.3. Two separately Fourier dualized function nodes are joined together to one large Fourier dualized function node.

The (\*)-part of Eq. (D.10) corresponds to the processing of the Fourier transformed messages according to the update rules for a summation-function node. This duality is shown in Fig. D.2.

### D.3 Dualization of Several Function Nodes

Consider the FFG shown in Fig. D.3(left). Dualizing them according to Secs. D.1 and D.2 one obtains the FFG in Fig. D.3(right). But Fourier transforming a message  $m(x)$  twice, we obtain the  $m(-x)$ , i.e., the same message but with the  $x$ -axis reversed. So two FG/FG

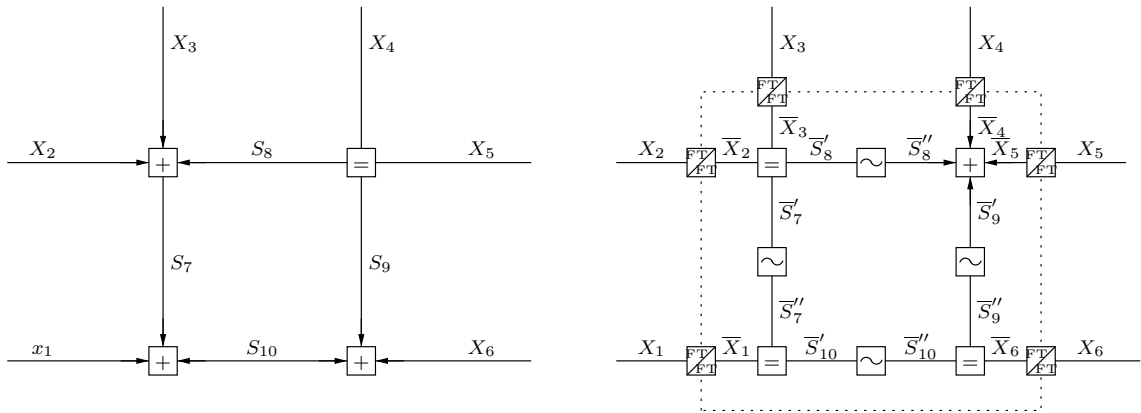


Figure D.5: Dualizing a loopy part of an FFG.

processing units can be replaced by one  $x$ -axis reversing unit as shown in Fig. D.4. Between  $\bar{s}'$  and  $\bar{s}''$  we have the relation:  $\bar{s}' = -\bar{s}''$ . It is obvious how this can be generalized to any subgraph which is a tree.

But this dualization also works for loopy graphs! Indeed, dualizing a loopy part of the FFG in Fig. D.5(left) gives Fig. D.5(right). It is dual in the sense that

$$\sum_{i=1}^6 x_i \bar{x}_i = 0 \quad (\text{D.11})$$

for all valid  $\mathbf{x}$  and all valid  $\bar{\mathbf{x}}$  configurations. (Note that the “internal” edges labelled  $s_j$ ,  $s'_j$ , and  $s''_j$  are not included in this sum.) That this indeed holds, can be seen from the following considerations. By duality of each function node alone we have the relations

$$x_1 \bar{x}_1 + s_7 \bar{s}_7'' + s_{10} \bar{s}_{10}' = 0, \quad (\text{D.12})$$

$$x_2 \bar{x}_2 + x_3 \bar{x}_3 + s_7 \bar{s}_7' + s_8 \bar{s}_8' = 0, \quad (\text{D.13})$$

$$x_4 \bar{x}_4 + x_5 \bar{x}_5 + s_8 \bar{s}_8'' + s_9 \bar{s}_9' = 0, \quad (\text{D.14})$$

$$x_6 \bar{x}_6 + s_9 \bar{s}_9'' + s_{10} \bar{s}_{10}'' = 0. \quad (\text{D.15})$$

Using these relations together with  $\bar{s}'_j + \bar{s}''_j = 0$  ( $j = 7, \dots, 10$ ) we obtain

$$\sum_{i=1}^6 x_i \bar{x}_i = \sum_{i=1}^6 x_i \bar{x}_i + \sum_{j=7}^{10} s_j (\bar{s}'_j + \bar{s}''_j) \quad (\text{D.16})$$

$$= (x_1 \bar{x}_1 + s_7 \bar{s}_7'' + s_{10} \bar{s}_{10}') + (x_2 \bar{x}_2 + x_3 \bar{x}_3 + s_7 \bar{s}_7' + s_8 \bar{s}_8') \quad (\text{D.17})$$

$$+ (x_4 \bar{x}_4 + x_5 \bar{x}_5 + s_8 \bar{s}_8'' + s_9 \bar{s}_9') + (x_6 \bar{x}_6 + s_9 \bar{s}_9'' + s_{10} \bar{s}_{10}'') \quad (\text{D.18})$$

$$= 0. \quad (\text{D.19})$$

From this example the principle how to prove the general case follows easily (see e.g. also [7]).

## Appendix E

# Planar Graph and Electrical Network Duality

For planar graphs, i.e. graphs that can be drawn on a plane without edges crossing each other, it is possible to define a dual graph. Planar electrical networks can be dualized in a similar way.<sup>1</sup>

### E.1 Planar Graphs

As stated above, planar graphs are graphs which can be drawn in the plane without having to cross the edges (see e.g. Fig. E.1(left)). Note that a graph is planar if and only if the graph can be drawn to the surface of a sphere (see p. 77 in [17]). This fact follows from using a stereographic projection of the sphere onto the plane. So basically we can identify the planar graphs with the polyhedrons for which we have Euler's formula

$$N_{\text{vertices}} + N_{\text{faces}} = N_{\text{edges}} + 2, \quad (\text{E.1})$$

where  $N_{\text{vertices}}$ ,  $N_{\text{faces}}$ , and  $N_{\text{edges}}$  denote the number of vertices, faces, and edges, respectively. This formula holds also when the graph is drawn in the plane: the number of faces is then

---

<sup>1</sup>In this appendix we closely follow [17].

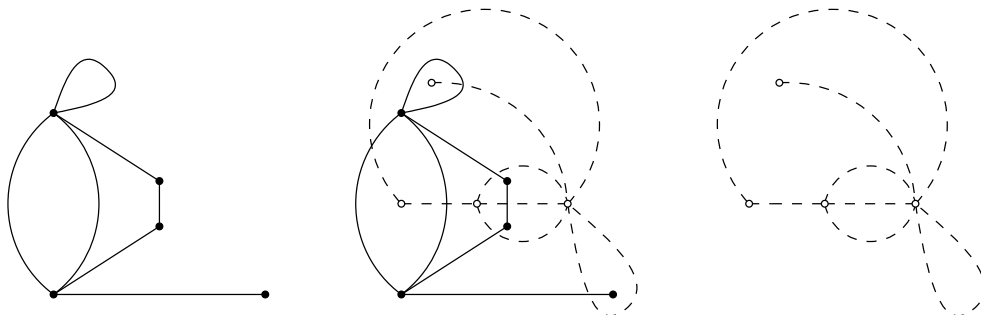


Figure E.1: Dualization of a graph. Left: original graph. Middle: original graph and dual graph. Right: dual graph. (Example taken from [17].)

the number of regions, where a region is either one of the bounded regions within the graph or the region around the graph. For a planar graph we therefore have

$$N_{\text{vertices}} + N_{\text{regions}} = N_{\text{edges}} + 2. \quad (\text{E.2})$$

E.g., for the graph in Fig. E.1(left) we have  $N_{\text{vertices}} = 5$ ,  $N_{\text{regions}} = 4$ ,  $N_{\text{edges}} = 7$ . The dual graph of a planar graph is then obtained in the following manner.

- Draw in each region of the graph a new vertex.
- There is a new edge between two new vertices if the regions corresponding to the new vertices are neighbors.
- The dual graph consists of the new vertices and the new edges.

This procedure is depicted in Figure E.1 for an exemplary planar graph. The dual graph is obviously again a planar graph and has the parameters

$$N_{\text{vertices}}^{\perp} = N_{\text{regions}}, \quad (\text{E.3})$$

$$N_{\text{regions}}^{\perp} = N_{\text{vertices}}, \quad (\text{E.4})$$

$$N_{\text{edges}}^{\perp} = N_{\text{edges}}, \quad (\text{E.5})$$

which fulfill

$$N_{\text{vertices}}^{\perp} + N_{\text{faces}}^{\perp} = N_{\text{edges}}^{\perp} + 2. \quad (\text{E.6})$$

The number of edges stayed the same because every new edge crosses an old edge and vice-versa. The meaning of vertices and regions have been interchanged in the process of dualization.

A word of caution is appropriate at this place. Graphs by themselves are defined abstractly and one and the same graph can have several graphical representations. Different graphical representations then of course represent isomorphic graphs. A problem arises when defining “the” dual of a graph. Dualizing different planar drawings of isomorphic graphs can lead to non-isomorphic graphs. Therefore, the concept of 2-isomorphism is introduced to basically identify all the possible dualizations of a graph.<sup>2</sup>

## E.2 Planar Electrical Networks

In the same way as planar graphs were defined, one can say that an electrical network is planar if it can be drawn without crossing the wires. As an example we consider the electrical network in Fig. E.2 which can be described by the system of equations (choosing  $(I_3, U_4)^T$  as state vector and  $(I_5, U_5)$  as outputs)

$$\begin{pmatrix} sL_3 I_3 \\ sC_4 U_4 \end{pmatrix} = \begin{pmatrix} -R_2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I_3 \\ U_4 \end{pmatrix} + \begin{pmatrix} 1 & -R_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U_1 \\ I_5 \end{pmatrix}, \quad (\text{E.7})$$

$$\begin{pmatrix} I_1 \\ U_5 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -R_2 & -1 \end{pmatrix} \begin{pmatrix} I_3 \\ U_4 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & -R_2 \end{pmatrix} \begin{pmatrix} U_1 \\ I_5 \end{pmatrix}. \quad (\text{E.8})$$

<sup>2</sup>Recall that two graphs are *isomorphic* if a one-one correspondence can be given both between their vertex sets and edge sets, preserving incidence. We call two graphs *2-isomorphic* if there is a one-one correspondence between their edge sets which preserves circuits. More exactly, a subset of edges form a circuit in one of the graphs if and only if the corresponding edges form a circuit (possibly in a different order) in the other graph. For more, see p. 72f. in [17].

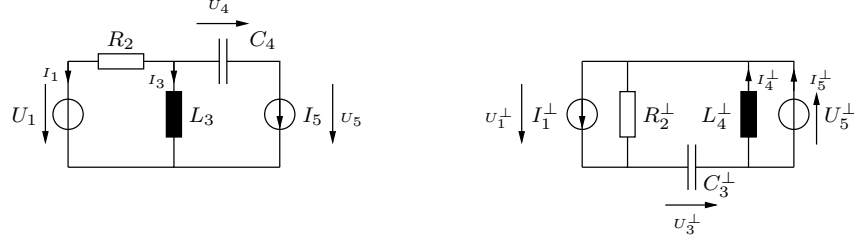


Figure E.2: Dualization of an electrical network: original electrical network, dual electrical network (Example taken from [17]).

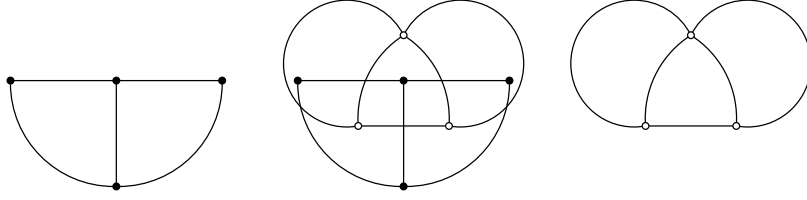


Figure E.3: Dualization of a underlying graph of the electrical network in Fig. E.2. Left: underlying graph of original electrical network. Middle: original graph and its dual. Right: dual graph (which is the underlying graph of the dual electrical network).

Replacing formally in Eqs. (E.7) and (E.8)

- $R_i$  by  $1/R_i^\perp$ , where  $R_i^\perp = 1/R_i$ ,
- $C_i$  by  $L_i^\perp$ , where  $L_i^\perp = C_i$ ,
- $L_i$  by  $C_i^\perp$ , where  $C_i^\perp = L_i$ ,
- $U_i$  by  $I_i^\perp$ , where  $I_i^\perp = U_i$ ,
- $I_i$  by  $U_i^\perp$ , where  $U_i^\perp = I_i$ ,

we obtain the description

$$\begin{pmatrix} sC_3^\perp U_3^\perp \\ sL_4^\perp I_4^\perp \end{pmatrix} = \begin{pmatrix} -1/R_2^\perp & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U_3^\perp \\ I_4^\perp \end{pmatrix} + \begin{pmatrix} 1 & -1/R_2^\perp \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_1^\perp \\ U_5^\perp \end{pmatrix} \quad (\text{E.9})$$

$$\begin{pmatrix} U_1^\perp \\ I_5^\perp \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -1/R_2^\perp & -1 \end{pmatrix} \begin{pmatrix} U_3^\perp \\ I_4^\perp \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & -1/R_2^\perp \end{pmatrix} \begin{pmatrix} I_1^\perp \\ U_5^\perp \end{pmatrix} \quad (\text{E.10})$$

of the dual electrical network as shown in Fig. E.2(right), see also Fig. E.3. This formal procedure is equivalent to the following graphical procedure.

- Dualize the underlying graph as done in Sec. E.1.
- There is a one-to-one mapping between old and new edges.
- Replace each resistor with value  $R$  on an old edge by a conductor of value  $R$  (or a resistor of value  $1/R$ ) on the corresponding new edge.

Voltage	Current
Resistance (impedance)	Admittance
Capacitance	Inductance
Short circuit	Open circuit
Parallel connection	Series connection
Voltage source	Current source
Capacitor	Inductor
Resistor	Resistor
Kirchhoff's Voltage Law	Kirchhoff's Current Law
Method of the "loop currents"	Method of the "node potentials"

Table E.1: Correspondence among network concepts related to the classical duality principle of electrical networks (taken from [17]).

- Replace each capacitor with value  $C$  on an old edge by a inductor of value  $C$  on the corresponding new edge.
- Replace each inductor with value  $L$  on an old edge by a capacitor of value  $L$  on the corresponding new edge.
- Replace a voltage source by a current source and vice-versa. (Care must be taken in choosing the correct direction of the voltage and the current.)

The procedure above is called *dualizing an electrical network*. It has the correspondences shown in Table E.1.

## Appendix F

# Singular Value Decomposition

### F.1 The Singular Value Decomposition

In this section we follow the resumee given in [8]. A square complex matrix  $\mathbf{A}$  is *unitary* if  $\mathbf{A}\mathbf{A}^H = \mathbf{1}$ , i.e.,  $\mathbf{A}^{-1} = \mathbf{A}^H$ . Any matrix  $\mathbf{A}$  (not necessarily square) can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^H, \quad (\text{F.1})$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are unitary matrices, and where the matrix  $\mathbf{S}$  (not necessarily square) is real and diagonal with positive diagonal entries only:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (\text{F.2})$$

with  $\mathbf{S}' := \text{diag}(\sigma_1, \dots, \sigma_\ell)$ ,  $\sigma_k > 0$ ,  $k = 1, \dots, \ell$ . The decomposition in Eq. (F.1) is called the *singular value decomposition (SVD)* of  $\mathbf{A}$ .

### F.2 The Moore-Penrose Generalized Inverse

The *Moore-Penrose generalized inverse* or *pseudo-inverse* of a complex matrix  $\mathbf{A}$  (not necessarily square) with singular value decomposition (F.1) is the matrix

$$\mathbf{A}^\# := \mathbf{V}\mathbf{S}^\#\mathbf{U}^H \quad (\text{F.3})$$

with

$$\mathbf{S}^\# := \begin{pmatrix} \text{diag}(\sigma_1^{-1}, \dots, \sigma_\ell^{-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \quad (\text{F.4})$$

Note that  $\mathbf{S}^\#$  has the same dimension as  $\mathbf{S}^H$  and  $\mathbf{A}^\#$  has the same dimension as  $\mathbf{A}^H$ . It is easy to see that

$$\mathbf{A}\mathbf{A}^\#\mathbf{A} = \mathbf{A}, \quad (\text{F.5})$$

$$\mathbf{A}^\#\mathbf{A}\mathbf{A}^\# = \mathbf{A}^\#, \quad (\text{F.6})$$

and

$$\mathbf{A}^\#\mathbf{x} = \mathbf{0} \iff \mathbf{x}^H\mathbf{A} = \mathbf{0}. \quad (\text{F.7})$$

It follows from Eqs. (F.5) and (F.7) that  $\mathbf{A}\mathbf{A}^H$  is the projection onto the column-space of  $\mathbf{A}$ . If the rank of  $\mathbf{A}$  equals the number of rows, then  $\mathbf{A}\mathbf{A}^\# = \mathbb{1}$  and

$$\mathbf{A}^\# = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}; \quad (\text{F.8})$$

if the rank of  $\mathbf{A}$  equals the number of columns, then  $\mathbf{A}^\#\mathbf{A} = \mathbb{1}$  and

$$\mathbf{A}^\# = (\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}^H. \quad (\text{F.9})$$

Another useful formula is

$$(\mathbf{A}\mathbf{A}^H)^\# = (\mathbf{A}^\#)^H\mathbf{A}^\#. \quad (\text{F.10})$$



# Appendix G

## Tables

(Tables are given on the next few pages.)

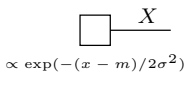
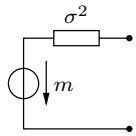

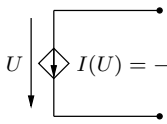
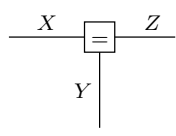
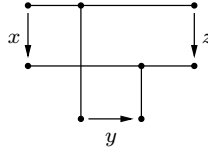
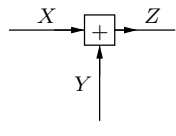
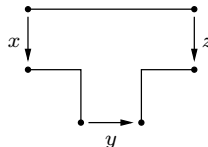
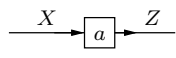
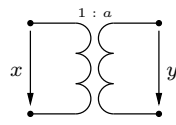
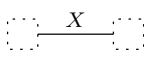
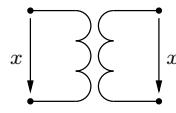
1	 $\propto \exp(-(x - m)/2\sigma^2)$	
2	 $f(\cdot)$	 $I(U) = -\partial \ln f(x) / \partial x _{x=U}$
3		
4		
5		
6		

Table G.1: Correspondence between the nodes and edges in the factor graph (left) and the components of the electrical network (right). All variables are scalars.

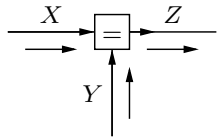
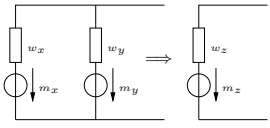
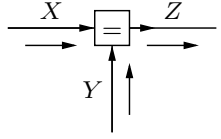
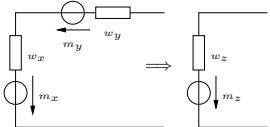
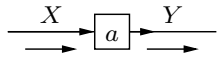
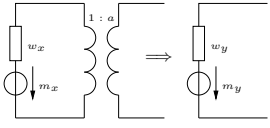
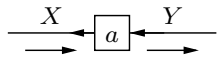
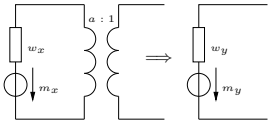
1		$m_Z = (w_X + w_Y)^{-1}(w_X m_X + w_Y m_Y)$ $w_Z = w_X + w_Y$ $v_Z = (v_X + v_Y)^{-1} v_X v_Y$	$\begin{pmatrix} 1 & 0 & 0 \\ w_Y & 1 & -m_Y w_Y \\ 0 & 0 & 1 \end{pmatrix}$	
2		$m_Z = m_X + m_Y$ $v_Z = v_X + v_Y$ $w_Z = (w_X + w_Y)^{-1} w_X w_Y$	$\begin{pmatrix} 1 & v_Y & m_Y \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
3		$m_y = a m_x$ $v_y = a^2 v_x$	$\begin{pmatrix} a & 0 & 0 \\ 0 & a^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
4		$m_y = a^{-1} m_x$ $w_y = a^2 w_x$	$\begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix}$	

Table G.2: Computation of messages in the case of jointly Gaussian random variables (first three columns; for interpretation of third column, see Sec. 4.3) with circuit interpretation (fourth column).

Distribution function	Probability density function	Name of Element	$I$ - $U$ characteristic
Normal distribution	$p_X(x) \propto \exp(-(x - m)/\sigma^2)$	Linear resistor	$I(U) = \frac{1}{\sigma^2} \cdot (U - m)$
Laplace distribution	$p_X(x) \propto \exp(-\lambda x - m )$	Sign resistor	$I(U) = \lambda \operatorname{sign}(U - m)$
Generalized logistic distribution	$p_X(x) \propto \cosh(\lambda_1(x - m))^{-\lambda_2}$	tanh resistor	$I(U) = \lambda_1 \lambda_2 \tanh(\lambda_1(x - m))$
“Diode distribution”	$p_X(x) \propto \exp(-I_0(U_T \exp(x/U_T) - x))$	Diode	$I(U) = I_0 (\exp(x/U_T) - 1)$

Table G.3: Dictionary of corresponding pdfs and  $I(U)$ -characteristics

# Bibliography

- [1] J. B. Dennis, *Mathematical Programming and Electrical Networks*. The Technology Press of The Massachusetts Institute of Technology, Cambridge, Mass., 1959.
- [2] R. C. Davis and H.-A. Loeliger, “A Nonalgorithmic Maximum Likelihood Decoder for Trellis Codes,” *IEEE Trans. on Inform. Theory*, vol. IT-39, no. 4, pp. 1450–1453, 1993.
- [3] H.-A. Loeliger, F. Lustenberger, M. Helfenstein, and F. Tarköy, “Probability Propagation and Decoding in Analog VLSI,” *IEEE Trans. on Inform. Theory*, vol. IT-47, no. 2, pp. 837–843, 2001.
- [4] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [5] D. W. Carter, “A Circuit Theory of the Kalman Filter,” in *Proceedings of the 32nd Conference on Decision and Control*, (San Antonio, TX, USA), pp. 1224–1226, Dec. 1993.
- [6] D. Lippuner, Nov. 2001. private communication.
- [7] G. D. Forney, Jr., “Codes on Graphs: Normal Realizations,” *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 520–548, 2001.
- [8] H.-A. Loeliger, “Least Squares and Kalman Filtering on Forney Graphs,” *preprint for Forney Festschrift*, 2001.
- [9] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor Graphs and the Sum-Product Algorithm,” *IEEE Trans. on Inform. Theory*, vol. IT-47, no. 2, pp. 498–519, 2001.
- [10] J. C. Willems, “Paradigms and Puzzles in the Theory of Dynamical Systems,” *IEEE Trans. Automat. Control*, vol. 36, no. 3, pp. 259–294, 1991.
- [11] N. Wiberg, *Codes and Decoding on General Graphs*. PhD thesis, Linköping University, Sweden, 1996.
- [12] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, N.J.: Prentice-Hall, 1979.
- [13] Y. Weiss and W. T. Freeman, “On the Optimality of the Max-Product Belief Propagation Algorithm in Arbitrary Graphs,” *IEEE Trans. on Inform. Theory*, vol. IT-47, no. 2, pp. 736–744, 2001.
- [14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

- 
- [15] T. Kailath, *Linear Systems*. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1980. Prentice-Hall Information and System Sciences Series.
  - [16] P. Bamberg and S. Sternberg, *A Course in Mathematics for Students of Physics. 2*. Cambridge: Cambridge University Press, 1991.
  - [17] A. Recski, *Matroid Theory and its Applications in Electric Network Theory and in Statistics*. Berlin: Springer-Verlag, 1989.
  - [18] Y. Mao and F. R. Kschischang, “On Factor Graphs and the Fourier Transform,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, (Washington, D.C., USA), p. 224, June 24-29 2001.
  - [19] S. Boyd and L. Vandenberghe, “Convex optimization,” *preprint of book*, 2001.
  - [20] T. L. Magnanti, “Fenchel and Lagrange Duality are Equivalent,” *Math. Programming*, vol. 7, pp. 253–258, 1974.