Representations and Updating of Gaussian Messages in Factor Graphs

Pascal O. Vontobel*

April 9, 2003

Abstract

We discuss different possibilities to represent Gaussian messages in a factor graph. One of them, called the projective message represenation, can handle very general situations: for this resentation we derive the update rules for indicator function nodes which represent linear constraints. Compared to [1] and [2], we do *not* need singular value decompositions (SVDs) and pseudo-inverses.

1 Introduction

We consider Forney-style factor graphs (FFGs), see e.g. [3] (there called normal graphs) or [1]. Fig. 1 depicts the situation we are interested in. Based on the incoming messages $\mu_X(x)$ and $\mu_Y(y)$, we would like to compute the outgoing message $\mu_Z(z)$ according to the sum-product algorithm [4]. In this paper we focus on the important case

- where $f(x, y, z)$ is an indicator function whose valid configurations of X, Y, and Z are given by a linear constraint and
- where the imcoming messages $\mu_X(x)$ and $\mu_Y(y)$ are Gaussian densities. This is e.g. the case when the global function is a jointly Gaussian density, or simply when the messages are set to be Gaussian.

Our aim is to show an alternative to the approaches shown in [1] and [2]. Note that in this paper we only present the case where edge variables are real scalars or real vectors; it is not difficult to generalize the results to the complex case (using proper complex random variables [5]) as considered in [1].

2 Different Representations of Gaussian Messages in the Strict Sense

Definition 1 *We call a message $\mu_X(x)$ a* Gaussian message in the strict sense *when it can be written as*

$$\mu_X(x) = \mathrm{e}^{-\phi_X(x)} \tag{1}$$

*Coordinated Science Lab, University of Illinois at Urbana-Champaign, IL, USA, `vontobel@ifp.uiuc.edu`.
Supported by NSF Grant CCR 99-84515.

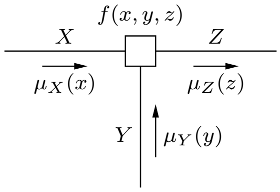


Figure 1: FFG with function node $f(x,y,z)$ and (half-)edges X, Y, and Z.

with

$$\phi_X(x) \triangleq \frac{1}{2}(x - m_X)^{\mathsf{T}} W_X (x - m_X) + c, \tag{2}$$

and where W_X *is an* $n_X \times n_X$ *positive definite matrix, i.e.,* $\phi(x)$ *is a quadratic function in* x*. Two message that differ only by a scale factor will be considered to be equivalent, i.e., we do not require a normalization of the message such that* $\int_{-\infty}^{+\infty} \mu_X(x)\, dx = 1$*.*

The origin of the terminology "Gaussian message" is fairly obvious: up to a multiplicative constant such a message is the density of a Gaussian random variable. As is well known, a Gaussian density is entirely defined by giving the covariance matrix V_X and the mean m_X. Obviously we have $V_X \triangleq W_X^{-1}$.[1] It will turn out to be useful to define the "modified mean vector" $w_X \triangleq W_X m_X$. The following nomenclature (i.e. moments/canonical parameters) can also be found e.g. in [6].

Definition 2 *A Gaussian message in the strict sense can be defined either by*

- *the moments* (V_X, m_X)*, or*
- *the canonical parameters* (W_X, w_X)*.*

To switch between the pairs (V_X, m_X) *and* (W_X, w_X) *we can apply the rules*

$$(V_X, m_X) \mapsto (W_X, w_X) = (V_X^{-1}, V_X^{-1} m_X), \tag{3}$$

$$(W_X, w_X) \mapsto (V_X, m_X) = (W_X^{-1}, W_X^{-1} w_X). \tag{4}$$

(Observe the symmetrical expressions in (3) and (4).)

Definition 3 *Given a message* $\mu_X(x)$ *(not necessarily Gaussian), we can define the (negative)* score function *(see e.g. [7])*

$$s_X : \mathbb{R}^{n_X} \to \mathbb{R}^{n_X} \tag{5}$$

$$x \mapsto \nabla_x \phi_X(x) = -\frac{\nabla_x \mu_X(x)}{\mu_X(x)} = -\nabla_x \log\left(\mu_X(x)\right), \tag{6}$$

where ∇_x *denotes the gradient with respect to* x*.*[2] *Note, that given a score function* $s_X(x)$*, we can find (up to a multiplicative constant) the corresponding message.*

[1] Because W_X is assumed to be a positive definite matrix, it is invertible. It follows that in this case V_X is also positive definite.

[2] The gradient ∇_x of a scalar function with respect to x yields a column vector function.

In the following, we assume $\mu_X(x)$ to be a Gaussian message in the strict sense as in (1) and (2). Then we get the score function

$$s_X(x) = W_X(x - m_X) = W_X x - W_X m_X = W_X x - w_X. \tag{7}$$

Lemma 4 *The score function of a Gaussian message in the strict sense is a linear and bijective function from $\mathbb{R}^{n_X}$ to $\mathbb{R}^{n_X}$. (From the bijectivity follows also the invertibility of the score function of a Gaussian message in the strict sense.)*

Proof: This follows easily from (7) and the invertibility of W_X. □

Because $s_X(\cdot)$ is a linear mapping from $\mathbb{R}^{n_X}$ to $\mathbb{R}^{n_X}$, there exist $n_X \times n_X$-matrices $T_{X,1}$ and $T_{X,2}$ and a column vector $t_{X,3}$ of length n_X such that

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \cdot \begin{pmatrix} x \\ s_X(x) \\ 1 \end{pmatrix} = 0 \tag{8}$$

for all x. The result (7) indicates, that we can e.g. choose[3]

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -W_X & \mathbb{1} & w_X \end{pmatrix}. \tag{9}$$

Of course, multiplying $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}$ from the left by an invertible matrix yields another suitable triple $(T_{X,1}, T_{X,2}, t_{X,3})$, i.e. such a transformation does not change the equation system (8). E.g. multiplying from the left by $W_X^{-1} = V_X$ gives

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X & m_X \end{pmatrix}. \tag{10}$$

It is not difficult to see that all valid pairs $(x, s_X(x))$ form an n_X-dimensional *affine* space in a $2n_X$-dimensional space. An interpretation as an n_X-dimensional subspace of a $2n_X$-dimensional *projective* space is also possible: one considers all $(x, \tilde{x}, x')$ that are orthogonal to $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}$.

Definition 5 *We call a triple $(T_{X,1}, T_{X,2}, t_{X,3})$ that defines a Gaussian message in the strict sense a* projective message representation *(of a Gaussian message). Sometimes we will also use the matrix $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}$ to represent the triple.*

To conclude this subsection, we list the different possibilities for representing a Gaussian message in the strict sense:

- by the function $\mu_X(\cdot)$,
- by the function $\phi_X(\cdot)$,
- by the function $s_X(\cdot)$,
- by the pair (V_X, m_X),
- by the pair (W_X, w_X),

[3]The matrix $\mathbb{1}$ denotes an identity matrix of a suitable size.

- by a triple $(T_{X,1}, T_{X,2}, t_{X,3})$.

While $\mu_X(\cdot)$, $\phi_X(\cdot)$, $s_X(\cdot)$ give all the information we need to know, they cannot easily be stored as they are functions with domain $\mathbb{R}^{n_X}$. But as we have seen, it is entirely sufficient to store either (V_X, m_X), (W_X, w_X) or $(T_{X,1}, T_{X,2}, t_{X,3})$ in order to recover the message function.

3 Update of Gaussian Messages in the Strict Sense

We now consider the update of messages according to the sum-product algorithm for indicator function nodes that represent linear constraints. We discuss only an indicator function node with three edges attached to it (as e.g. shown in Fig. 1); from this example, the general principle can easily be derived. We assume the incoming messages to be Gaussian messages in the strict sense represented by the projective message representation.

3.1 Indicator Function Node

So, assume to have an indicator function node $f(x, y, z)$ as in Fig. 1 with the three edges X, Y, and Z attached to it. Furthermore, assume that the function can be written as[4] $f(x, y, z) = \delta(A_X x + A_Y y + A_Z z)$ with matrices A_X, A_Y, and A_Z such that $\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix}$ has full rank. The triple $(x^\mathsf{T}, y^\mathsf{T}, z^\mathsf{T})^\mathsf{T}$ is called a valid configuration if the function evaluates to a nonzero value, i.e. if and only if

$$A_X x + A_Y y + A_Z z = 0 \qquad \Longleftrightarrow \qquad \begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0. \tag{11}$$

Moreover, we will need some matrices A'_X, A'_Y, A'_Z defining the dual behavior such that the rows of $\begin{pmatrix} A'_X & A'_Y & A'_Z \end{pmatrix}$ are orthogonal to the rows of $\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix}$ and such that the matrix

$$\begin{pmatrix} A_X & A_Y & A_Z \\ A'_X & A'_Y & A'_Z \end{pmatrix} \tag{12}$$

is a square matrix.

3.2 Updating the Messages

Theorem 6 *(Later on, this theorem will be shown to hold also for Gaussian messages in the wide sense, see Th. 11.) Consider an indicator function node as shown in Fig. 1 where the valid configurations are characterized by*

$$\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix}, \tag{13}$$

and its dual behavior is characterized by

$$\begin{pmatrix} A'_X & A'_Y & A'_Z \end{pmatrix}. \tag{14}$$

[4]When the argument is a scalar, the function $\delta(\cdot)$ is the usual Dirac delta function. When the argument is a vector, $\delta(\cdot)$ is the product of the Dirac delta function applied to each component separately.

Furthermore, assume that the incoming messages $\mu_X(x)$ and $\mu_Y(y)$ are Gaussian messages in the strict sense given in the projective message representation,

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \text{ and } \begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix}, \tag{15}$$

respectively. The outgoing message $\mu_Z(z)$ is a Gaussian message whose projective message representation can be computed by the following procedure. Form the matrix

$$\left(\begin{array}{ccccccc} T_{X,1} & T_{X,2} & 0 & 0 & 0 & 0 & t_{X,3} \\ 0 & 0 & T_{Y,1} & T_{Y,2} & 0 & 0 & t_{Y,3} \\ \hline A_X & 0 & A_Y & 0 & A_Z & 0 & 0 \\ \hline 0 & A'_X & 0 & A'_Y & 0 & -A'_Z & 0 \end{array}\right), \tag{16}$$

(note the minus sign in front of the matrix A'_Z) and apply to it some row transformations, i.e. multiply from left by an invertible matrix, in order to bring it into the row-reduced echelon form

$$\begin{pmatrix} \times & \times & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix}, \tag{17}$$

where the "$\times$"s are placeholders for partial matrices, and the second row of block matrices has n_Z lines. The entries $T_{Z,1}$, $T_{Z,2}$, and $t_{Z,3}$ yield the desired projective message representation of the outgoing message, viz.,

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix}. \tag{18}$$

(In Th. 11 this theorem is extended to to Gaussian messages in the wide sense.)

Proof: The incoming messages are

$$\mu_X(x) \triangleq \mathrm{e}^{-\phi_X(x)}, \qquad \mu_Y(y) \triangleq \mathrm{e}^{-\phi_Y(y)}. \tag{19}$$

where $\mu_X(\,\cdot\,)$ and $\mu_Y(\,\cdot\,)$ are defined by the projective message representations $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}$ and $\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix}$, respectively. Following the sum-product algorithm [4], we have

$$\begin{aligned} \mu_Z(z) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y,z)\mu_X(x)\mu_Y(y)\mathrm{d}y\,\mathrm{d}x & (20)\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y,z)\,\mathrm{e}^{-\phi_X(x)-\phi_Y(y)}\,\mathrm{d}y\,\mathrm{d}x. & (21) \end{aligned}$$

We introduce two sets, namely the set of all z where $\mu_Z(z)$ is not zero, and the set of all z such that there are valid configurations (x,y,z) of the function node under consideration, i.e.,

$$\begin{aligned} \mathcal{S}_Z^{(1)} &\triangleq \big\{ z \in \mathbb{R}^{n_Z} \,|\, \mu_Z(z) \neq 0 \big\}, \\ \mathcal{S}_Z^{(2)} &\triangleq \big\{ z \in \mathbb{R}^{n_Z} \,|\, \exists\,(x,y) \in \mathbb{R}^{n_X} \times \mathbb{R}^{n_Y} \text{ such that } A_X x + A_Y y + A_Z z = 0 \big\}. \end{aligned}$$

Because of the assumptions on $\mu_X(\,\cdot\,)$ and $\mu_Y(\,\cdot\,)$, the two sets are identical, i.e., $\mathcal{S}_Z^{(1)} = \mathcal{S}_Z^{(2)}$. For any $z \in \mathcal{S}_Z^{(2)}$ we have

$$\mu_Z(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y,z)\, \mathrm{e}^{-\phi_X(x)-\phi_Y(y)}\, \mathrm{d}y\, \mathrm{d}x \tag{22}$$

$$\overset{(*)}{\propto} \max_{\substack{x,y \\ A_X x + A_Y y + A_Z z = 0}} \mathrm{e}^{-\phi_X(x)-\phi_Y(y)} \tag{23}$$

$$= \exp\left(- \min_{\substack{x,y \\ A_X x + A_Y y + A_Z z = 0}} \phi_X(x) + \phi_Y(y) \right), \tag{24}$$

where step $(*)$ follows from the observation that the integral can be replaced by a maximization in the situation at hand (see e.g. Ch. 3 of [1]). Without loss of generality, the outgoing message $\mu_Z(z)$ can be written as

$$\mu_Z(z) \overset{\triangle}{=} \mathrm{e}^{-\phi_Z(z)}, \tag{25}$$

for some function $\phi_Z(z)$ for all $z \in \mathcal{S}_Z^{(1)}$; therefore

$$\phi_Z(z) = \left(\min_{\substack{x,y \\ A_X x + A_Y y + A_Z z = 0}} \phi_X(x) + \phi_Y(y) \right) + c \tag{26}$$

for some arbitrary constant c, which we assume for simplicity to be zero. For each z we have to find the minumum in (26). In order to do so, we formulate a Lagrangian (which is parametrized by z)

$$L(z) \overset{\triangle}{=} \phi_X(x) + \phi_Y(y) - \lambda(z)^{\mathsf{T}}(A_X x + A_Y y + A_Z z), \tag{27}$$

and set the gradient of this Lagrangian equal to zero, viz.,

$$0 = \nabla_x L(z)\big|_{x=x^*(z),y=y^*(z)} = \nabla_x \phi_X(x)\big|_{x=x^*(z),y=y^*(z)} - A_X^{\mathsf{T}}\lambda(z), \tag{28}$$

$$0 = \nabla_y L(z)\big|_{x=x^*(z),y=y^*(z)} = \nabla_y \phi_Y(y)\big|_{x=x^*(z),y=y^*(z)} - A_Y^{\mathsf{T}}\lambda(z), \tag{29}$$

$$0 = \nabla_{\lambda(z)} L(z)\big|_{x=x^*(z),y=y^*(z)} = -(A_X x^*(z) + A_Y y^*(z) + A_Z z). \tag{30}$$

The functions $x^*(z)$ and $y^*(z)$ represent the values of x and y where the minimum is achieved. To proceed, we need the score functions

$$s_X(x) \overset{\triangle}{=} \nabla_x \phi_X(x), \tag{31}$$

$$s_Y(y) \overset{\triangle}{=} \nabla_y \phi_Y(y), \tag{32}$$

$$s_Z(y) \overset{\triangle}{=} \nabla_z \phi_Z(z). \tag{33}$$

Because of the assumptions that the incoming messages are Gaussian messages in the strict sense, the score functions $s_X(x)$ and $s_Y(y)$ are invertible. Therefore, from (28) and (29), together with (31) and (32), we get

$$s_X(x^*(z)) = A_X^{\mathsf{T}}\lambda(z), \tag{34}$$

$$s_Y(y^*(z)) = A_Y^{\mathsf{T}}\lambda(z), \tag{35}$$

and by the invertibility of $s_X(x)$ and $s_Y(y)$ we see that for each z there is exactly one solution $(x^*(z), y^*(z))$. From (26) we therefore get

$$\phi_Z(z) = \phi_X(x^*(z)) + \phi_Y(y^*(z)). \tag{36}$$

Taking the gradient of this expression with respect to z we obtain[5]

$$\begin{aligned}
s_Z(z) &= \nabla_z \phi_Z(z) && (37)\\
&= \left(\nabla_z x^*(z)^\mathsf{T}\right) s_X(x^*(z)) + \left(\nabla_z y^*(z)^\mathsf{T}\right) s_Y(y^*(z)) && (38)\\
&\overset{(*)}{=} \left(\nabla_z x^*(z)^\mathsf{T}\right) A_X^\mathsf{T} \lambda(z) + \left(\nabla_z y^*(z)^\mathsf{T}\right) A_Y^\mathsf{T} \lambda(z) && (39)\\
&= \left[\left(\nabla_z x^*(z)^\mathsf{T}\right) A_X^\mathsf{T} + \left(\nabla_z y^*(z)^\mathsf{T}\right) A_Y^\mathsf{T}\right] \lambda(z), && (40)
\end{aligned}$$

where $(*)$ follows from (34) and (35). Transposing (30), viz,

$$x^*(z)^\mathsf{T} A_X^\mathsf{T} + y^*(z)^\mathsf{T} A_Y^\mathsf{T} + z^\mathsf{T} A_Z^\mathsf{T} = 0^\mathsf{T}, \tag{41}$$

and taking the gradient with respect to z yields

$$\left(\nabla_z x^*(z)^\mathsf{T}\right) A_X^\mathsf{T} + \left(\nabla_z y^*(z)^\mathsf{T}\right) A_Y^\mathsf{T} + A_Z^\mathsf{T} = 0. \tag{42}$$

Combining (40) and (42) we obtain

$$s_Z(z) = -A_Z^\mathsf{T} \lambda(z). \tag{43}$$

The important relations we have found so far, viz. (30), (34), (35), and (43), can be summarized by

$$\left\{ \begin{array}{rcl}
A_X x^*(z) + A_Y y^*(z) + A_Z z &=& 0,\\
s_X(x^*(z)) &=& A_X^\mathsf{T} \lambda(z)\\
s_Y(y^*(z)) &=& A_Y^\mathsf{T} \lambda(z)\\
s_Z(z) &=& -A_Z^\mathsf{T} \lambda(z)
\end{array} \right. \tag{44}$$

or equivalently by

$$\left\{ \begin{array}{rl}
\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix} \begin{pmatrix} x^*(z) \\ y^*(z) \\ z \end{pmatrix} & = 0\\
\begin{pmatrix} s_X(x^*(z))^\mathsf{T} & s_Y(y^*(z))^\mathsf{T} & s_Z(z)^\mathsf{T} \end{pmatrix} & = \lambda(z)^\mathsf{T} \cdot \begin{pmatrix} A_X & A_Y & -A_Z \end{pmatrix}
\end{array} \right. \tag{45}$$

Upon observing that $(A'_X, A'_Y, -A'_Z)$ spans the dual space of $(A_X, A_Y, -A_Z)$ (see Sec. 3.1), we see that (45) is equivalent to

$$\left\{ \begin{array}{rcl}
\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix} \begin{pmatrix} x^*(z) \\ y^*(z) \\ z \end{pmatrix} &=& 0\\
\begin{pmatrix} A'_X & A'_Y & -A'_Z \end{pmatrix} \begin{pmatrix} s_X(x^*(z)) \\ s_Y(y^*(z)) \\ s_Z(z) \end{pmatrix} &=& 0
\end{array} \right. \tag{46}$$

[5]Note that the gradient of a row vector function (as e.g. $x^*(z)^\mathsf{T}$) yields a matrix function.

which can be expressed by one big equation

$$\left(\begin{array}{ccccccc} A_X & 0 & A_Y & 0 & A_Z & 0 & 0 \\ \hline 0 & A'_X & 0 & A'_Y & 0 & -A'_Z & 0 \end{array}\right) \begin{pmatrix} x^*(z) \\ s_X(x^*(z)) \\ y^*(z) \\ s_Y(y^*(z)) \\ z \\ s_Z(z) \\ 1 \end{pmatrix} = 0. \tag{47}$$

Note the minus sign in front of the matrix the matrix A'_Z in (47). (The seventh column is a dummy column here, it will turn out to be useful as soon as we combine the above equation system with the message representations.)

Because we assumed in the theorem that the messages $\mu_X(\cdot)$ and $\mu_Y(\cdot)$ are given in the projective message representation, we can write for all z (see (8))

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \cdot \begin{pmatrix} x^*(z) \\ s_X(x^*(z)) \\ 1 \end{pmatrix} = 0, \quad \text{and} \tag{48}$$

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \cdot \begin{pmatrix} y^*(z) \\ s_Y(y^*(z)) \\ 1 \end{pmatrix} = 0, \tag{49}$$

respectively. Joining (47), (48), and (49) yields

$$\left(\begin{array}{ccccccc} T_{X,1} & T_{X,2} & 0 & 0 & 0 & 0 & t_{X,3} \\ 0 & 0 & T_{Y,1} & T_{Y,2} & 0 & 0 & t_{Y,3} \\ \hline A_X & 0 & A_Y & 0 & A_Z & 0 & 0 \\ \hline 0 & A'_X & 0 & A'_Y & 0 & -A'_Z & 0 \end{array}\right) \begin{pmatrix} x^*(z) \\ s_X(x^*(z)) \\ y^*(z) \\ s_Y(y^*(z)) \\ z \\ s_Z(z) \\ 1 \end{pmatrix} = 0. \tag{50}$$

Applying some row transformation, i.e. multiplying from left by any invertible matrix, in order to bring the equation system in (50) into the row-reduced echelon form

$$\begin{pmatrix} \times & \times & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix} \begin{pmatrix} x^*(z) \\ s_X(x^*(z)) \\ y^*(z) \\ s_Y(y^*(z)) \\ z \\ s_Z(z) \\ 1 \end{pmatrix} = 0, \tag{51}$$

where the "$\times$"s are placeholders for partial matrices, gives us the desired message representation for $\mu_Z(z)$, viz.,

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix} \cdot \begin{pmatrix} z \\ s_Z(z) \\ 1 \end{pmatrix} = 0, \tag{52}$$

(Note that the above process is essentially a projection onto the last three columns.)

□

Note: At this point the reader can already have a look at the examples in Sec. 5.

4 Representation of Gaussian Messages in the Wide Sense

Note: I guess the presentation in this chapter is now precise enough. But perhaps one can make still make in more precise.

4.1 Limitations of the Pairs (V_X, m_X) and (W_X, w_X)

We have seen in the last section that a Gaussian message in the strict sense can be characterized by the pair (V_X, m_X) or the pair (W_X, w_X). A good feature about these parametrizations is that their entries have a statistical meaning. The bad thing is that they are not general enough, i.e. in general we would like to allow the variance to be zero or infinity. If only variances smaller than infinity appear we can stick to the first pair, and if only variances larger than zero appear we can stick to the second pair. But neither pair is sufficient if we consider e.g. a random vector of length two with two independent components where the first component has variance zero and the second component has variance infinity.

More specifically, consider a message $\mu_X(x)$ along a branch $X = (X_1, X_2)^\mathsf{T}$, which is equal to the density of two jointly Gaussian random variables. Let the first random variable have zero mean and variance ε, let the second random variable have zero mean and variance $1/\varepsilon$, and let the two random variables be statistically independent. (Finally, we will be interested in the limit $\varepsilon \to 0$.) So this message can be represented by the pair

$$(V_X(\varepsilon), m_X(\varepsilon)) = \left(\begin{pmatrix} \varepsilon & 0 \\ 0 & 1/\varepsilon \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \tag{53}$$

or the pair

$$(W_X(\varepsilon), w_X(\varepsilon)) = \left(\begin{pmatrix} 1/\varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right). \tag{54}$$

But allowing only real entries in the matrices and vectors, we get *in both cases* a problem when we consider the limit $\varepsilon \to 0$, a case one is sometimes interested in.

These are the circumstances where the triple $(T_{X,1}, T_{X,2}, t_{X,3})$ turns out to be useful: it can not only represent the Gaussian messages in the strict sense (as shown in the last section), but it can also handle the more delicate cases (one of which has just been mentioned in the last paragraph).

In the case of messages where $n_X = 1$, the problem is the well-known problem of different parametrization of straight lines in a two-dimensional plane. To be more specific, let us assume the plane to be the set of all points $(x, y) \in \mathbb{R}^2$. A straight line can be written as the set of all points $(x, y) \in \mathbb{R}^2$ where $y = \alpha_1 x + \beta_1$. All straight lines can be represented in this fashion except those which are parallel to the y-axis. Similarly, a straight line can be written as the set of all points $(x, y) \in \mathbb{R}^2$ where $x = \alpha_2 y + \beta_2$. With this parametrization, all straight lines can be described except those that are parallel to the x-axis. A solution to this dilemma is the parametrization $Ax + By + C = 0$ with which all straight lines can be represented. Note that a straight line with this parametrization can be seen as given by the intersection of the planes $Ax + By + Cz = 0$ and $z = 1$ in $\mathbb{R}^3$.

4.2 Gaussian Messages in the Wide Sense

To motivate what comes next, let us offer the following analogy, namely how in algebra the set of the real numbers $\mathbb{R}$ is usually defined based on the set of rational numbers $\mathbb{Q}$. First, we define the metric $d(x, y) = |x - y|$ on the set $\mathbb{Q}$. Let a Cauchy sequence in $\mathbb{Q}$ be a sequence $x_1, x_2, \ldots$ with $x_k \in \mathbb{Q}$ such that $\lim_{\min(m,n)\to\infty} d(x_m, x_n) = 0$. It is easy to see that every rational number x is the limit of a Cauchy sequence, namely e.g. the sequence where $x_k = x$ for all $k \geq 1$. We now introduce a new set of objects, namely the set of all equivalence classes of Caucy sequences in $\mathbb{Q}$, where we identify two Cauchy sequences if their difference has the limit point 0. By extending addition and multiplication from $\mathbb{Q}$ to this set, we obtain an algebraic object which is isomorphic to the field of reals $\mathbb{R}$.

The generalization of messages from Gaussian messages in the strict sense to Gaussian messages in the wide sense will be along the same lines (although the above analogy is not to be taken word-for-word). To the set of rational numbers $\mathbb{Q}$ will correspond the set of Gaussian messages in the strict sense. Given a distance measure on this set, we can consider Cauchy sequences of messages. To the set of reals $\mathbb{R}$, defined above as the set of equivalence classes of Cauchy sequences, will then correspond the set of Gaussian messages in the wide sense.

To apply this procedure, we must first define a distance between two messages; this will be done after the next definition.

Definition 7 *To the paramtrization $(T_{X,1}, T_{X,2}, t_{X,3})$ of a Gaussian message we can associate the following n_X-dimensional affine subspace in a $2n_X$-dimensional space, namely the set $\mathcal{X}$ of all points $(x^\mathsf{T}, \tilde{x}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{2n_X}$ such that*

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \begin{pmatrix} x \\ \tilde{x} \\ 1 \end{pmatrix} = 0. \tag{55}$$

Note that this subspace is well-defined as equivalent $(T_{X,1}, T_{X,2}, t_{X,3})$ messages, i.e. messages whose matrices $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}$ are equal up to multiplication from the left by an invertible matrix, yield the same subspace. Equivalently, to the paramtrization $(T_{X,1}, T_{X,2}, t_{X,3})$ of a Gaussian message we can associate an n_X-dimensional projective subspace in a $2n_X$-dimensional projective space, namely the set $\mathcal{X}$ of all points $(x^\mathsf{T}, \tilde{x}^\mathsf{T}, x')^\mathsf{T} \in \mathbb{R}^{2n_X+1}$ such that

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \begin{pmatrix} x \\ \tilde{x} \\ x' \end{pmatrix} = 0. \tag{56}$$

(Or as an $n_X + 1$-dimensional subspace in an $2n_X + 1$-dimensional affine space.)

Definition 8 *We define the distance between two messages in the following way. As defined in Def. 7 we construct for each message its (projective) subspace; the distance is then given by the gap metric (see App. A).*

Given such a distance measure, we can define Cauchy sequences of Gaussian messages in the strict sense. E.g., assuming that the sequence $\varepsilon_1, \varepsilon_2, \ldots$ converges to zero, the sequence

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}(\varepsilon_k) \triangleq \begin{pmatrix} -\mathbb{1} & V_X(\varepsilon_k) & m_X(\varepsilon_k) \end{pmatrix} \tag{57}$$

$$\triangleq \begin{pmatrix} -1 & 0 & \varepsilon_k & 0 & 0 \\ 0 & -1 & 0 & 1/\varepsilon_k & 0 \end{pmatrix}. \tag{58}$$

for $k \geq 1$ is a Cauchy sequency of Gaussian messages in the strict sense. Using the fact that we can multiply from the left by the invertible matrix $\begin{pmatrix} 1 & 0 \\ 0 & \varepsilon_k \end{pmatrix}$ without changing the message that is being represented (and also without changing the subspaces defined in Def. 7) , we get

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}(\varepsilon_k) \propto \begin{pmatrix} -1 & 0 & \varepsilon_k & 0 & 0 \\ 0 & -\varepsilon_k & 0 & 1 & 0 \end{pmatrix}. \tag{59}$$

Now it is easy to see that in the limit $k \to \infty$ we get

$$\lim_{k \to \infty} \begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}(\varepsilon_k) = \begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}(0) \propto \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{60}$$

This is now a projective message representation for the above case where the first variance is zero, whereas the second variance is infinite. Note that whereas all messages $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}(\varepsilon_k)$ were Gaussian messages in the strict sense, $\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix}(0)$ is not!

Definition 9 *A* Gaussian message in the wide sense *is a message that can be represented as the limit of a Cauchy sequence of Gaussian messages in the strict sense.*

For such a Gaussian message in the wide sense that is given in the projective message representation $\begin{pmatrix} T_{X,1} & T_{X,2} & T_{X,2} \end{pmatrix}$, the submatrix $\begin{pmatrix} T_{X,1} & T_{X,2} \end{pmatrix}$ must have full rank.

Given a Gaussian message in the wide sense, there are infinitely many Caucy sequences with that limit. The next lemma answers the question, if it is important which Cauchy sequence one takes.

Lemma 10 *Let the incoming messages to an indicator function node be defined by Cauchy sequences. In the cases where the combination of the messages with the help of the sum-product algorithm "is well-posed", it does not matter what Cauchy sequences are actually chosen.*

Theorem 11 *The procedure in Th. 6 can be extended to hold also for Gaussian messages in the wide sense for those cases where the combination of the messages "is well-posed"; see also Lemma 10. Note that one can calculate directly with the limit of the Cauchy sequence that represents a Gaussian message in the wide sense.*

Basically, one can derive general rules when an "well-posed" situation occurs: the matrix $\begin{pmatrix} T_{Z,1} & T_{Z,2} \end{pmatrix}$ *obtained from (17) in Th. 6 must have full rank.*

4.3 An Example From Electrical Networks

To illustrate what we mean by "being well-posed" in Lemma 10, let us give an example using electrical networks [8]. In [8] we showed that one can associate an FFG to an electrical network. The simplification of the electrical network can then be interpreted as performing the max-product algorithm on the FFG.

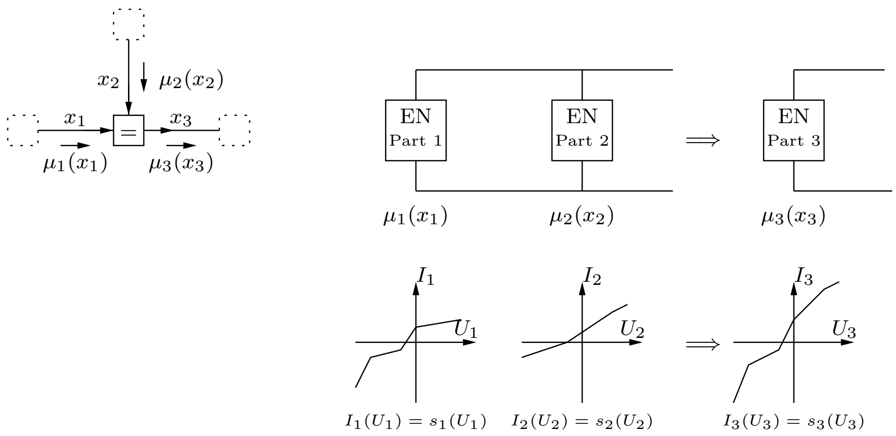


Figure 2: Electrical network interpretation of the max-product update rule for an equal function node. (Note that in the case of linear networks the current-voltage characteristics are linear. Furthermore note that electrical network components with nonlinear current-voltage characteristic are combined using the max-product algorithm; but in the case of linear elements, the max-product algorithm is equivalent to the sum-product algorithm.)

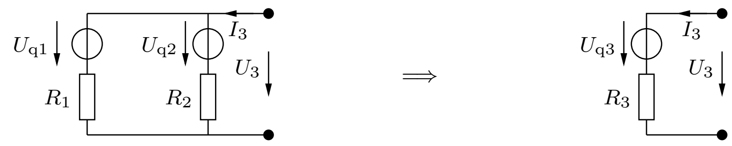


Figure 3: Electrical network interpretation of the max-product/sum-product update rule for an equal function node.

In the case of linear electrical networks, the max-product algorithm is equivalent to the sum-product algorithm and the messages of the voltage-mode FFG correspond to real voltage sources $U = U_\mathrm{q} + R_\mathrm{q} I$ (i.e. an ideal voltage source in series with a linear resistor), where the ideal voltage source U_q equals the mean and the linear resistor R_q equals the variance of the Gaussian message, respectively. (There is also a current-mode FFG and an equivalent interpretation with real current sources.)

Consider Figs. 2 and 3. The equivalent of the processing of two Gaussian messages in the FFG is the replacement of two parallel real voltage sources (i.e. $U = U_\mathrm{q1} + R_\mathrm{q1} I$ and $U = U_\mathrm{q2} + R_\mathrm{q2} I$) by one real voltage source $U = U_\mathrm{q3} + R_\mathrm{q3} I$ with

$$U_\mathrm{q3} = \frac{\frac{U_\mathrm{q1}}{R_\mathrm{q1}} + \frac{U_\mathrm{q2}}{R_\mathrm{q2}}}{\frac{1}{R_\mathrm{q1}} + \frac{1}{R_\mathrm{q2}}}, \tag{61}$$

$$R_\mathrm{q3} = \frac{1}{\frac{1}{R_\mathrm{q1}} + \frac{1}{R_\mathrm{q2}}}. \tag{62}$$

We consider now the following sequences of $\{U_{\mathrm{q1},k}\}$, $\{R_{\mathrm{q1},k}\}$, $\{U_{\mathrm{q2},k}\}$, and $\{R_{\mathrm{q2},k}\}$: $U_{\mathrm{q1},k} = U_\mathrm{q1}$ for all $k \geq 0$, $\lim_{k\to\infty} R_{\mathrm{q1},k} = 0$, $U_{\mathrm{q2},k} = U_\mathrm{q2}$ for all $k \geq 0$, and $\lim_{k\to\infty} R_{\mathrm{q2},k} = 0$. Independently on the convergence behavior of $\{R_{\mathrm{q1},k}\}$ and $\{R_{\mathrm{q2},k}\}$, we always have $\lim_{k\to\infty} R_{\mathrm{q3},k} = 0$. But unless $\lim_{k\to\infty}(U_{\mathrm{q1},k} - U_{\mathrm{q2},k}) = 0$, the limit $\lim_{k\to\infty} U_{\mathrm{q3},k}$ strongly depends on the convergence behavior of $\{R_{\mathrm{q1},k}\}$ and $\{R_{\mathrm{q2},k}\}$!

Translated into FFGs and messages, the above scenario corresponds in the limit to the case where two Gaussian messages entering a degree-three equal node have each variance zero. Combining these two messages makes only sense in the very rare case that both means are equal. In the usual case, where the messages have different means, one cannot combine the two message. In this sense the above "being well-posed" has to be understood.

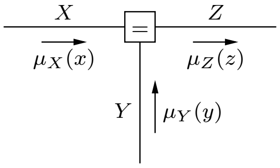
4.4 A Little Duality Result

We conclude this section with a little duality result.

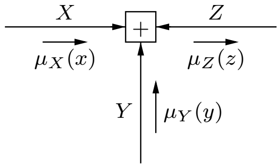
Lemma 12 *Given the scheme (16) for some indicator function node and some incoming messages, the scheme (16) for the dual indicator function node and the same incoming messages is obtained by exchanging* (A_X, A_Y, A_Z) *and* (A'_X, A'_Y, A'_Z) *in the former scheme.*

5 Examples of Message Updates

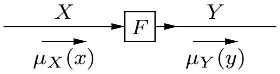
The following sections feature different examples which show how to use the projective message representation to update messages. The examples correspond one-to-one to the entries 1 to 6 in Tab. 1 of [1]. Example 7 is not related to an example in [1].



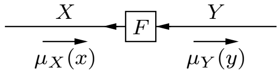
(a) Example 1



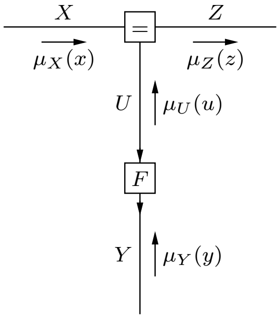
(b) Example 2



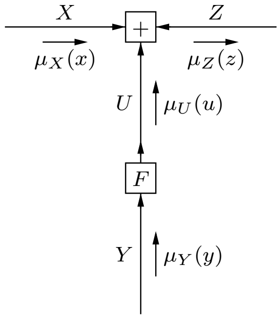
(c) Example 3



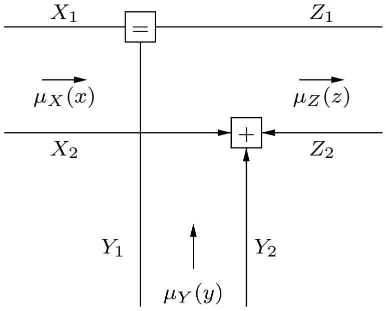
(d) Example 4



(e) Example 5



(f) Example 6



(g) Example 7

Figure 4: The different FFGs corresponding to the different examples treated in this section.

5.1 Example 1: Equality Indicator Function Node

We assume to have an indicator function node representing the function $X = Y = Z$ as in Fig. 4(a), viz.,

$$\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix} \triangleq \begin{pmatrix} \mathbb{1} & 0 & -\mathbb{1} \\ 0 & \mathbb{1} & -\mathbb{1} \end{pmatrix}. \tag{63}$$

The dual constraint is

$$\begin{pmatrix} A'_X & A'_Y & A'_Z \end{pmatrix} \triangleq \begin{pmatrix} \mathbb{1} & \mathbb{1} & \mathbb{1} \end{pmatrix}. \tag{64}$$

The matrix corresponding to (16) is

$$\left(\begin{array}{ccccccc} T_{X,1} & T_{X,2} & 0 & 0 & 0 & 0 & t_{X,3} \\ 0 & 0 & T_{Y,1} & T_{Y,2} & 0 & 0 & t_{Y,3} \\ \hline \mathbb{1} & 0 & 0 & 0 & -\mathbb{1} & 0 & 0 \\ 0 & 0 & \mathbb{1} & 0 & -\mathbb{1} & 0 & 0 \\ \hline 0 & \mathbb{1} & 0 & \mathbb{1} & 0 & -\mathbb{1} & 0 \end{array}\right). \tag{65}$$

To be specific, we assume the incoming messages to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -W_X & \mathbb{1} & w_X \end{pmatrix}, \tag{66}$$

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} -W_Y & \mathbb{1} & w_Y \end{pmatrix}. \tag{67}$$

Therefore, (65) turns into

$$\left(\begin{array}{ccccccc} -W_X & \mathbb{1} & 0 & 0 & 0 & 0 & w_X \\ 0 & 0 & -W_Y & \mathbb{1} & 0 & 0 & w_Y \\ \hline \mathbb{1} & 0 & 0 & 0 & -\mathbb{1} & 0 & 0 \\ 0 & 0 & \mathbb{1} & 0 & -\mathbb{1} & 0 & 0 \\ \hline 0 & \mathbb{1} & 0 & \mathbb{1} & 0 & -\mathbb{1} & 0 \end{array}\right). \tag{68}$$

We add the following to the first row: $\mathbb{1}$ times the second row, W_X times the third row, W_Y times the fourth row, and $-\mathbb{1}$ times the fifth row; we obtain

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -(W_X + W_Y) & \mathbb{1} & w_X + w_Y \end{pmatrix}. \tag{69}$$

This yields the outgoing message

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix} \triangleq \begin{pmatrix} -W_Z & \mathbb{1} & w_Z \end{pmatrix} \triangleq \begin{pmatrix} -(W_X + W_Y) & \mathbb{1} & w_X + w_Y \end{pmatrix}. \tag{70}$$

5.2 Example 2: Summation Indicator Function Node

We assume to have an indicator function node representing the function $X + Y + Z = 0$ as in Fig. 4(b), viz.,

$$\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix} \triangleq \begin{pmatrix} \mathbb{1} & \mathbb{1} & \mathbb{1} \end{pmatrix}. \tag{71}$$

The dual constraint is

$$\begin{pmatrix} A'_X & A'_Y & A'_Z \end{pmatrix} \triangleq \begin{pmatrix} \mathbb{1} & 0 & -\mathbb{1} \\ 0 & \mathbb{1} & -\mathbb{1} \end{pmatrix}. \tag{72}$$

The matrix corresponding to (16) is

$$\left(\begin{array}{ccccccc} T_{X,1} & T_{X,2} & 0 & 0 & 0 & 0 & t_{X,3} \\ 0 & 0 & T_{Y,1} & T_{Y,2} & 0 & 0 & t_{Y,3} \\ \hline \mathbb{1} & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 & 0 \\ \hline 0 & \mathbb{1} & 0 & 0 & 0 & \mathbb{1} & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 \end{array}\right). \tag{73}$$

To be specific, we assume the incoming messages to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X & m_X \end{pmatrix}, \tag{74}$$

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_Y & m_Y \end{pmatrix}. \tag{75}$$

Therefore, (73) turns into

$$\left(\begin{array}{ccccccc} -\mathbb{1} & V_X & 0 & 0 & 0 & 0 & m_X \\ 0 & 0 & -\mathbb{1} & V_Y & 0 & 0 & m_Y \\ \hline \mathbb{1} & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 & 0 \\ \hline 0 & \mathbb{1} & 0 & 0 & 0 & \mathbb{1} & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 \end{array}\right). \tag{76}$$

We add the following to the first row: $\mathbb{1}$ times the second row, $\mathbb{1}$ times the third row, $-V_X$ times the fourth row, and $-V_Y$ times the fifth row; we obtain

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \mathbb{1} & -(V_X + V_Y) & m_X + m_Y \end{pmatrix}. \tag{77}$$

This yields the outgoing message

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & R_Z & m_Z \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X + V_Y & -(m_X + m_Y) \end{pmatrix}. \tag{78}$$

5.3 Example 3: Multiplication by a Matrix (Forward Direction)

We assume to have an indicator function node representing the function $Y = F \cdot X$ as in Fig. 4(c), viz.,

$$\begin{pmatrix} A_X & A_Y \end{pmatrix} \triangleq \begin{pmatrix} F & -\mathbb{1} \end{pmatrix}. \tag{79}$$

The dual constraint is

$$\begin{pmatrix} A'_X & A'_Y \end{pmatrix} \triangleq \begin{pmatrix} \mathbb{1} & F^\mathsf{T} \end{pmatrix}. \tag{80}$$

The matrix corresponding to (16) is

$$\left(\begin{array}{ccccc} T_{X,1} & T_{X,2} & 0 & 0 & t_{X,3} \\ \hline F & 0 & -\mathbb{1} & 0 & 0 \\ \hline 0 & \mathbb{1} & 0 & -F^\mathsf{T} & 0 \end{array}\right). \tag{81}$$

To be specific, we assume the incoming message to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X & m_X \end{pmatrix}. \tag{82}$$

Therefore, (81) turns into

$$\left(\begin{array}{ccccc} -\mathbb{1} & V_X & 0 & 0 & m_X \\ \hline F & 0 & -\mathbb{1} & 0 & 0 \\ \hline 0 & \mathbb{1} & 0 & -F^\mathsf{T} & 0 \end{array}\right). \tag{83}$$

We add the following to the second row: F times the first row and $-FV_X$ times the third row; we obtain

$$\begin{pmatrix} 0 & 0 & -\mathbb{1} & FV_XF^\mathsf{T} & Fm_X \end{pmatrix}. \tag{84}$$

This yields the outgoing message

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_Y & m_Y \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & FV_XF^\mathsf{T} & Fm_X \end{pmatrix}. \tag{85}$$

5.4 Example 4: Multiplication by a Matrix (Reverse Direction)

We assume to have an indicator function node representing the function $X = F \cdot Y$ as in Fig. 4(d), viz.,

$$\begin{pmatrix} A_X & A_Y \end{pmatrix} \triangleq \begin{pmatrix} \mathbb{1} & -F \end{pmatrix}. \tag{86}$$

The dual constraint is

$$\begin{pmatrix} A'_X & A'_Y \end{pmatrix} \triangleq \begin{pmatrix} F^{\mathsf{T}} & \mathbb{1} \end{pmatrix}. \tag{87}$$

The matrix corresponding to (16) is

$$\left(\begin{array}{ccccc} T_{X,1} & T_{X,2} & 0 & 0 & t_{X,3} \\ \hline \mathbb{1} & 0 & -F & 0 & 0 \\ \hline 0 & F^{\mathsf{T}} & 0 & -\mathbb{1} & 0 \end{array}\right). \tag{88}$$

To be specific, we assume the incoming message to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -W_X & \mathbb{1} & w_X \end{pmatrix}. \tag{89}$$

Therefore, (88) turns into

$$\left(\begin{array}{ccccc} -W_X & \mathbb{1} & 0 & 0 & w_X \\ \hline \mathbb{1} & 0 & -F & 0 & 0 \\ \hline 0 & F^{\mathsf{T}} & 0 & -\mathbb{1} & 0 \end{array}\right). \tag{90}$$

We add the following to the third row: $-F^{\mathsf{T}}$ times the first row and $-F^{\mathsf{T}} W_X$ times the second row; we obtain

$$\begin{pmatrix} 0 & 0 & F^{\mathsf{T}} W_X F & -\mathbb{1} & -F^{\mathsf{T}} w_X \end{pmatrix}. \tag{91}$$

This yields the outgoing message

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} -W_Y & \mathbb{1} & w_Y \end{pmatrix} \triangleq \begin{pmatrix} -F^{\mathsf{T}} W_X F & \mathbb{1} & F^{\mathsf{T}} w_X \end{pmatrix}. \tag{92}$$

5.5 Example 5: Equality Function Node and Multiplication by a Matrix

We assume to have indicator function nodes representing the functions $X = U = Z$ and $Y = FU$ as in Fig. 4(e). (We replaced the W in [1] by U.) To be specfic, we assume the incoming messages to be

$$\left(T_{X,1} \quad T_{X,2} \quad t_{X,3}\right) \triangleq \left(-W_X \quad \mathbb{1} \quad w_X\right), \tag{93}$$

$$\left(T_{Y,1} \quad T_{Y,2} \quad t_{Y,3}\right) \triangleq \left(-W_Y \quad \mathbb{1} \quad w_Y\right). \tag{94}$$

According to Sec. 5.4, the intermediate message is

$$\left(T_{U,1} \quad T_{U,2} \quad t_{U,3}\right) \triangleq \left(-W_U \quad \mathbb{1} \quad w_U\right) \triangleq \left(-F^{\mathsf{T}} W_Y F \quad \mathbb{1} \quad F^{\mathsf{T}} w_Y\right). \tag{95}$$

Using the results of Sec. 5.1 we get the outgoing message

$$\left(T_{Z,1} \quad T_{Z,2} \quad t_{Z,3}\right) \triangleq \left(-W_Z \quad \mathbb{1} \quad w_Z\right) \tag{96}$$

$$\triangleq \left(-\left(W_X + F^{\mathsf{T}} W_Y F\right) \quad \mathbb{1} \quad w_X + F^{\mathsf{T}} w_Y\right). \tag{97}$$

Assuming W_X and W_Y to be invertible such that $V_X = W_X^{-1}$ and $V_Y = W_Y^{-1}$ we can reformulate the outgoing message to

$$\left(T_{Z,1} \quad T_{Z,2} \quad t_{Z,3}\right) \propto \left(-\mathbb{1} \quad V_Z \quad m_Z\right) \tag{98}$$

$$\triangleq \left(-\mathbb{1} \quad V_X - V_X F^{\mathsf{T}} G F V_X \quad m_X + V_X F^{\mathsf{T}} G (m_Y - F m_X)\right), \tag{99}$$

where

$$G \triangleq \left(F V_X F^{\mathsf{T}} + V_Y\right)^{-1}. \tag{100}$$

This follows by the matrix inversion lemma, viz.,

$$(H_1 + H_2 H_3 H_4)^{-1} = H_1^{-1} - H_1^{-1} H_2 \left(H_4 H_1^{-1} H_2 + H_3^{-1}\right)^{-1} H_4 H_1^{-1}, \tag{101}$$

from which we have

$$\left(W_X + F^{\mathsf{T}} W_Y F\right)^{-1} = V_X - V_X F^{\mathsf{T}} G F V_X \tag{102}$$

and

$$\left(W_X + F^{\mathsf{T}} W_Y F\right)^{-1} F^{\mathsf{T}} = \left(W_X + F^{\mathsf{T}} W_Y F\right)^{-1} F^{\mathsf{T}} (G V_Y)^{-1} (G V_Y) \tag{103}$$

$$= \left(W_X + F^{\mathsf{T}} W_Y F\right)^{-1} F^{\mathsf{T}} W_Y \left(F V_X F^{\mathsf{T}} + V_Y\right) G V_Y \tag{104}$$

$$= \left(W_X + F^{\mathsf{T}} W_Y F\right)^{-1} \left(F^{\mathsf{T}} W_Y F V_X F^{\mathsf{T}} + F^{\mathsf{T}}\right) G V_Y \tag{105}$$

$$= \left(W_X + F^{\mathsf{T}} W_Y F\right)^{-1} \left(F^{\mathsf{T}} W_Y F + W_X\right) V_X F^{\mathsf{T}} G V_Y \tag{106}$$

$$= V_X F^{\mathsf{T}} G V_Y. \tag{107}$$

5.6 Example 6: Summation Function Node and Multiplication by a Matrix

We assume to have an indicator function node representing the functions $X + U + Z = 0$ and $U = FY$ as in Fig. 4(f). (We replaced the W in [1] by U.) To be specific, we assume the incoming messages to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X & m_X \end{pmatrix}, \tag{108}$$

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_Y & m_Y \end{pmatrix}. \tag{109}$$

According to Sec. 5.3, the intermediate message is

$$\begin{pmatrix} T_{U,1} & T_{U,2} & t_{U,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_U & m_U \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & FV_YF^\mathsf{T} & Fm_Y \end{pmatrix}. \tag{110}$$

Using the results of Sec. 5.2 we get the outgoing message

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_Z & m_Z \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X + FV_YF^\mathsf{T} & -m_X - Fm_Y \end{pmatrix}. \tag{111}$$

Assuming V_X and V_Y to be invertible such that $W_X = V_X^{-1}$ and $W_Y = V_Y^{-1}$ we can reformulate the outgoing message to

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix} \propto \begin{pmatrix} -W_Z & \mathbb{1} & w_Z \end{pmatrix} \tag{112}$$

$$\triangleq \begin{pmatrix} -\left(W_X - W_XFGF^\mathsf{T}W_X\right) & \mathbb{1} & -\left(w_X + W_XFG(w_Y - F^\mathsf{T}w_X)\right) \end{pmatrix}. \tag{113}$$

with

$$G \triangleq \left(F^\mathsf{T}W_XF + W_Y\right)^{-1}. \tag{114}$$

This follows by the matrix inversion lemma, viz.,

$$(H_1 + H_2H_3H_4)^{-1} = H_1^{-1} - H_1^{-1}H_2\left(H_4H_1^{-1}H_2 + H_3^{-1}\right)^{-1}H_4H_1^{-1}, \tag{115}$$

from which we have

$$\left(V_X + FV_YF^\mathsf{T}\right)^{-1} = W_X - W_XFGF^\mathsf{T}W_X \tag{116}$$

and

$$\left(V_X + FV_YF^\mathsf{T}\right)^{-1} F = \left(V_X + FV_YF^\mathsf{T}\right)^{-1} F\,(GW_Y)^{-1}\,(GW_Y) \tag{117}$$

$$= \left(V_X + FV_YF^\mathsf{T}\right)^{-1} FV_Y\left(F^\mathsf{T}W_XF + W_Y\right)GW_Y \tag{118}$$

$$= \left(V_X + FV_YF^\mathsf{T}\right)^{-1}\left(FV_YF^\mathsf{T}W_XF + F\right)GW_Y \tag{119}$$

$$= \left(V_X + FV_YF^\mathsf{T}\right)^{-1}\left(FV_YF^\mathsf{T} + V_X\right)W_XFGW_Y \tag{120}$$

$$= W_XFGW_Y. \tag{121}$$

5.7 Example 7: Parallel Equality and Summation Function Nodes

We assume to have an indicator function node with $n_X = n_Y = n_Z = 2$ that represents $X_1 = Y_1 = Z_1$ and $X_2 + Y_2 + Z_2 = 0$ as in Fig. 4(g), viz.,

$$\begin{pmatrix} A_X & A_Y & A_Z \end{pmatrix} \triangleq \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{122}$$

The dual constraint is

$$\begin{pmatrix} A'_X & A'_Y & A'_Z \end{pmatrix} \triangleq \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}. \tag{123}$$

We assume the incoming messages to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} T^{11}_{X,1} & T^{12}_{X,1} & T^{11}_{X,2} & T^{12}_{X,2} & t^{1}_{X,3} \\ T^{21}_{X,1} & T^{22}_{X,1} & T^{21}_{X,2} & T^{22}_{X,2} & t^{2}_{X,3} \end{pmatrix}, \tag{124}$$

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} T^{11}_{Y,1} & T^{12}_{Y,1} & T^{11}_{Y,2} & T^{12}_{Y,2} & t^{1}_{Y,3} \\ T^{21}_{Y,1} & T^{22}_{Y,1} & T^{21}_{Y,2} & T^{22}_{Y,2} & t^{2}_{Y,3} \end{pmatrix}. \tag{125}$$

The matrix corresponding to (16) is

$$\left(\begin{array}{ccccccccccccc} T^{11}_{X,1} & T^{12}_{X,1} & T^{11}_{X,2} & T^{12}_{X,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t^{1}_{X,3} \\ T^{21}_{X,1} & T^{22}_{X,1} & T^{21}_{X,2} & T^{22}_{X,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t^{2}_{X,3} \\ 0 & 0 & 0 & 0 & T^{11}_{Y,1} & T^{12}_{Y,1} & T^{11}_{Y,2} & T^{12}_{Y,2} & 0 & 0 & 0 & 0 & t^{1}_{Y,3} \\ 0 & 0 & 0 & 0 & T^{21}_{Y,1} & T^{22}_{Y,1} & T^{21}_{Y,2} & T^{22}_{Y,2} & 0 & 0 & 0 & 0 & t^{2}_{Y,3} \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right). \tag{126}$$

From this we can obtain the outgoing message

$$\begin{pmatrix} T_{Z,1} & T_{Z,2} & t_{Z,3} \end{pmatrix}. \tag{127}$$

5.8 Example 8: Summation Function Node and Multiplication by a Matrix (Example 6 revisited)

We assume to have an indicator function node representing the function $X+U+Z=0$ and $U=FY$ as in Fig. 4(f). (We replaced the W in [1] by U.) To be specific, we assume the incoming messages to be

$$\begin{pmatrix} T_{X,1} & T_{X,2} & t_{X,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_X & m_X \end{pmatrix}, \tag{128}$$

$$\begin{pmatrix} T_{Y,1} & T_{Y,2} & t_{Y,3} \end{pmatrix} \triangleq \begin{pmatrix} -\mathbb{1} & V_Y & m_Y \end{pmatrix}. \tag{129}$$

$$\left(\begin{array}{ccccccccc} -W_X & \mathbb{1} & 0 & 0 & 0 & 0 & 0 & 0 & w_X \\ 0 & 0 & -W_Y & \mathbb{1} & 0 & 0 & 0 & 0 & w_Y \\ \hline 0 & 0 & F & 0 & -\mathbb{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 & -F^{\mathsf{T}} & 0 & 0 & 0 \\ \hline \mathbb{1} & 0 & 0 & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 & 0 \\ 0 & \mathbb{1} & 0 & 0 & 0 & 0 & 0 & \mathbb{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 \end{array}\right). \tag{130}$$

The first two columns correspond to the message along the X-edge, the next two columns correspond to the message along the Y-edge, the next two columns correspond to the message along the U-edge, the next two columns correspond to the message along the Z-edge, and finally the last column corresponds to (modified) means. In order to obtain the outgoing message we first derive from (130) a new scheme (with fewer rows) where the columns corresponding to the message along the U-edge are zero. So, multiplying (130) from the left by

$$\begin{pmatrix} \mathbb{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbb{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbb{1} & 0 & \mathbb{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 & 0 & F^{\mathsf{T}} \\ 0 & 0 & 0 & 0 & 0 & \mathbb{1} & 0 \end{pmatrix}. \tag{131}$$

we obtain

$$\begin{pmatrix} -W_X & \mathbb{1} & 0 & 0 & 0 & 0 & 0 & 0 & w_X \\ 0 & 0 & -W_Y & \mathbb{1} & 0 & 0 & 0 & 0 & w_Y \\ \mathbb{1} & 0 & F & 0 & 0 & 0 & \mathbb{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 & 0 & 0 & F^{\mathsf{T}} & 0 \\ 0 & \mathbb{1} & 0 & 0 & 0 & 0 & 0 & \mathbb{1} & 0 \end{pmatrix}. \tag{132}$$

Multiplying (132) from the left by

$$\begin{pmatrix} \mathbb{1} & 0 & 0 & 0 & 0 \\ 0 & -\mathbb{1} & F^{\mathsf{T}} W_X & 0 & 0 \\ 0 & 0 & \mathbb{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 \\ 0 & 0 & 0 & 0 & \mathbb{1} \end{pmatrix}. \tag{133}$$

we obtain

$$\begin{pmatrix} -W_X & \mathbb{1} & 0 & 0 & 0 & 0 & 0 & 0 & w_X \\ F^\mathsf{T} W_X & 0 & G^{-1} & -\mathbb{1} & 0 & 0 & F^\mathsf{T} W_X & 0 & -w_Y \\ \mathbb{1} & 0 & F & 0 & 0 & 0 & \mathbb{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1} & 0 & 0 & 0 & F^\mathsf{T} & 0 \\ 0 & \mathbb{1} & 0 & 0 & 0 & 0 & 0 & \mathbb{1} & 0 \end{pmatrix}. \tag{134}$$

where

$$G \triangleq \left(F^\mathsf{T} W_X F + W_Y\right)^{-1}. \tag{135}$$

Multiplying (134) from left with

$$\begin{pmatrix} -\left(\mathbb{1} - W_X F G F^\mathsf{T}\right) & W_X F G & -W_X & W_X F G & \left(\mathbb{1} - W_X F G F^\mathsf{T}\right) \end{pmatrix}. \tag{136}$$

we obtain

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -\left(W_X - W_X F G F^\mathsf{T} W_X\right) & \mathbb{1} & -\left(w_X + W_X F G (w_Y - F^\mathsf{T} w_X)\right) \end{pmatrix}. \tag{137}$$

6 Further topics

- Geometric interpretation of update rules. In the equal node case, the update operation corresponds to summing two planes in a certain direction. Similarly, in the sum node case, the update operation corresponds to summing two planes in some different direction. In the scaling node case, the update operation corresponds to scaling the axis. We will say more about that. Meanwhile, see [8].
- Grassman varieties and Schubert calculus. There is a connection to Grassmann varieties and also to Schubert calculus.
- Square-root algorithm. The framework presented here can be extended to include square-root algorithms known from Kalman-filtering; perhaps we will say something about that at a later time.

A Distances between Subspaces

To measure the "distance" between two subspaces one can use the gap metric [9]. Let P_1 and P_2 be the the projection matrix onto the first and second subspaces, respectively. The distance given by the gap metric is then defined to be

$$g(P_1, P_2) \triangleq ||P_1 - P_2|| \tag{138}$$

where $||\cdot||$ is some matrix norm.

Let a subspace be defined by $\{x \in \mathbb{R}^n : Mx = 0\}$, where M has full rank. Then the projection operator onto this subspace is $P_M = \mathbb{1} - M^\mathsf{T}(MM^\mathsf{T})^{-1}M$. The distance between two subspaces defined by M_1 and M_2, respectively, is then

$$g(M_1, M_2) = ||M_2^\mathsf{T}(M_2M_2^\mathsf{T})^{-1}M_2 - M_1^\mathsf{T}(M_1M_1^\mathsf{T})^{-1}M_1|| \tag{139}$$

(note that the two $\mathbb{1}$ matrices cancel each other). As a special case, consider two one-dimensional subspaces and the L_2 matrix norm: it can be shown that in this case the distance between the two subspaces equals the sine of the angle between the two subspaces.

Acknowledgments

I would like to thank Joachim Rosenthal from Notre Dame University for helping me in finding literature about metrics for subspaces and his comments about Grassmann varieties and Schubert calculus.

References

[1] H.-A. Loeliger, "Least squares and Kalman filtering on Forney graphs," in *Codes, Graphs, and Systems* (R. E. Blahut and R. Koetter, eds.), Kluwer, 2002. Festschrift in honor of G. D. Forney, Jr.

[2] G. D. Forney, Jr., "Notes on Gaussian random vectors, Fourier transforms of Gaussian distributions, and belief propagation," *preprint*, 2002.

[3] G. D. Forney, Jr., "Codes on graphs: normal realizations," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 520–548, 2001.

[4] 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.

[5] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. IT-39, no. 4, pp. 1293–1302, 1993.

[6] M. I. Jordan and C. M. Bishop, "An introduction to graphical models," *preprint*, 2000.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[8] P. O. Vontobel, *Kalman Filters, Factor Graphs, and Electrical Networks.* Post-Diploma Project, ETH Zurich, 2002. Available online under `http://www.isi.ee.ethz.ch/publications`.

[9] T. Kato, *Perturbation Theory for Linear Operators.* Berlin: Springer-Verlag, second ed., 1976. Grundlehren der Mathematischen Wissenschaften, Band 132.