Factor Graphs, Electrical Networks, and Entropy

Pascal O. Vontobel
Signal and Information Proc. Lab.
ETH Zurich
CH-8092 Zurich, Switzerland
vontobel@isi.ee.ethz.ch

Abstract

The aim of this paper is to highlight some connections between factor graphs, electrical networks, and differential entropy.

1 Introduction

At first sight, the three topics mentioned in the title seem to be unrelated and not to have too much in common. In this paper we would like to show that this is not so and that there are indeed connections. We will keep the discussion at a high level; for more mathematical details we encourage the reader to consult the referenced books and papers.

Some historical background on factor graphs and graphical models in general can e.g. be found in [1, 2] or [3]. We will, in fact, not use factor graphs as defined in [2], but a variation introduced by Forney [4] (there called "normal graphs"). The advantages of these Forney graphs (here called "Forney-style factor graphs" or shortly FFGs) were discussed in [2] and [5].

The facts we need from the theory of electrical networks can be found e.g. in [6], and the book by Cover and Thomas [7] contains essentially everything we need to know about differential entropies as far as concerns this paper. We will only talk about scalar random variables, but the results can be generalized to vector random variables.

In Sec. 2 we review some possible ways of going from FFGs to electrical networks and back. Sec. 3 presents some new results connecting results from electrical network theory with results about differential entropies, and finally in Sec. 4 we give some conclusions.

2 Connections between FFGs and Electrical Networks

There are different ways to derive the various relationships between FFGs and electrical networks. To give a flavor, we will in the following just point out two of them by giving specific examples, first going from an FFG to an electrical network, then going from an electrical network to FFGs. The notation that we use is explained in detail in [6, 3]. Most of the results in this section were originally motivated by the book by Dennis [8].

Example 1 [From an FFG to an electrical network, "voltage-mode derivation"] Fig. 1 shows an 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:

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

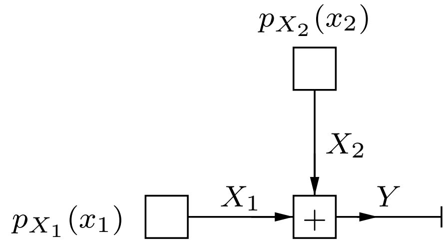


Figure 1: FFG for Ex. 1.

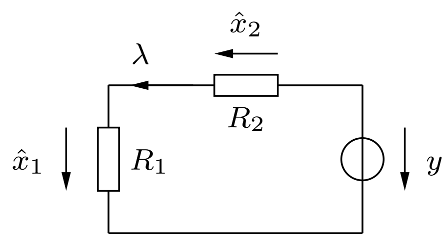


Figure 2: Electrical network for Ex. 1.

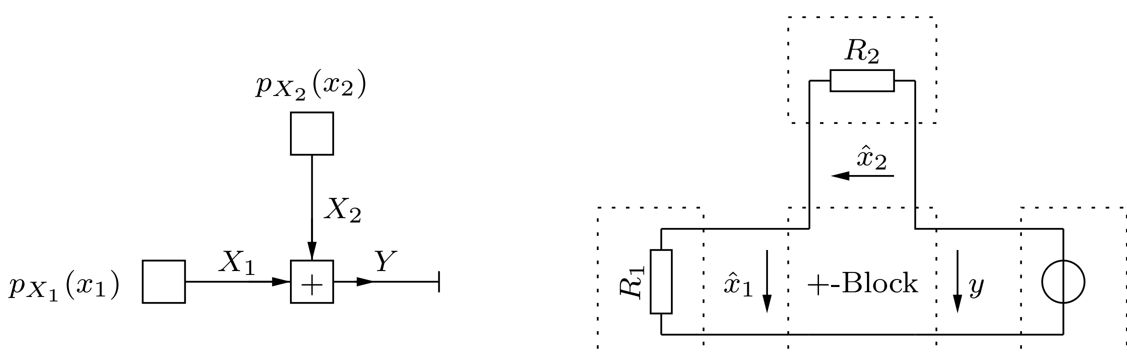


Figure 3: FFG versus electrical network in the case of Ex. 1.

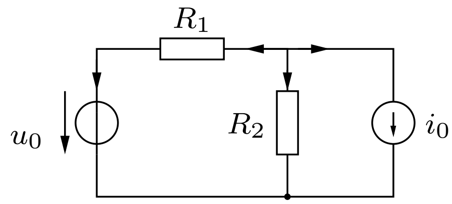


Figure 4: Electrical network for Ex. 2.

where $\delta(.)$ denotes the Dirac-delta distribution[1]. To be specific, we assume $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(0, \sigma_2^2)$:

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

Based on a measurement $Y = y$, we would like to find the blockwise MAP estimate $(\hat{x}_1, \hat{x}_2)$ of the vector (x_1, x_2), i.e.,

$$(\hat{x}_1, \hat{x}_2) = \arg \max_{(x_1, x_2)} p_{X_1 X_2 | Y}(x_1, x_2 | y) = \arg \max_{(x_1, x_2)} p_{X_1 X_2 Y}(x_1, x_2, y).$$

(Note that we maximize jointly!) 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 minimizing $-\ln p_{X_1}(x_1) - \ln p_{X_2}(x_2)$ under the constraint $y = x_1 + x_2$. The corresponding Lagrangian is

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

Setting the gradient of L equal to zero we obtain

$$\left\{ \begin{array}{rlll} \frac{\partial}{\partial x_1} L & = \frac{x_1}{\sigma_1^2} - \lambda & \stackrel{!}{=} 0 & \text{(component law)} \\ \frac{\partial}{\partial x_2} L & = \frac{x_2}{\sigma_2^2} - \lambda & \stackrel{!}{=} 0 & \text{(component law)} \\ \frac{\partial}{\partial \lambda} L & = y - x_1 - x_2 & \stackrel{!}{=} 0 & \text{(Kirchhoff voltage law)} \end{array} \right. \tag{1}$$

The electrical network in Fig. 2 implements the three equations from (1) with $R_1 = \sigma_1^2$ and $R_2 = \sigma_2^2$. The first two equations correspond to component law 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. 3 the electrical network is redrawn so as to show its close topological relationship to its FFG. The Lagrange multiplier λ turns out to be a current and loosely speaking plays the role of "exchanging information" from one part of the circuit to the other.

Ch. 2 in [6] gives more details about this example, where also other examples are given, Ch. 3 in[6] discusses the topological equivalence between FFGs and the corresponding electrical networks more generally, whereas Sec. 5.1 in [6] discusses generalizations to non-Gaussian distributions. $\triangle$

Example 2 [From an electrical network to an FFG] Dennis [8] looked at what optimization problems can be solved with the help of electrical networks. The key idea is the following. Given an electrical network, the solution (i.e. the branch voltages and

[1]We consider the Dirac-delta distribution to be the limes of a Gaussian density, see e.g. [3, 6].

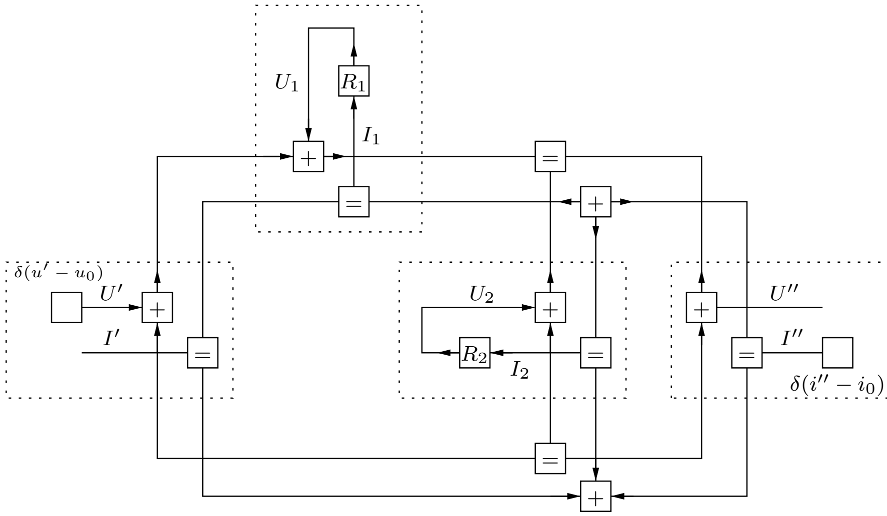


Figure 5: FFG (voltage/current version) corresponding to the electrical network in Fig. 4.

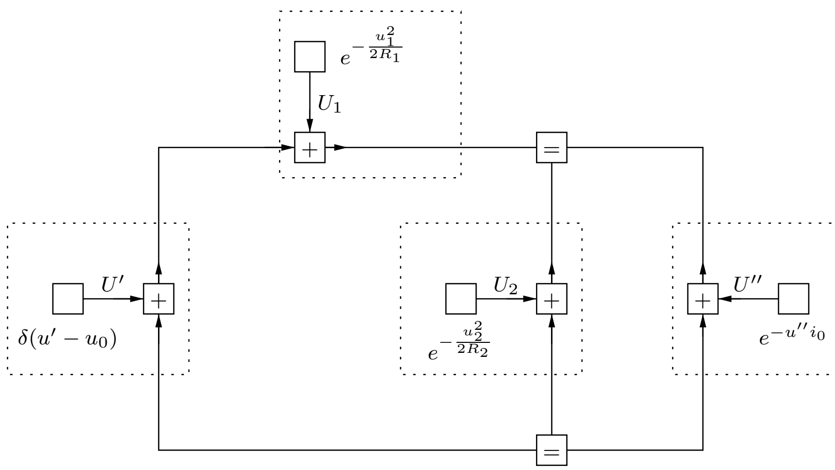


Figure 6: FFG (voltage version) corresponding to the electrical network in Fig. 4.

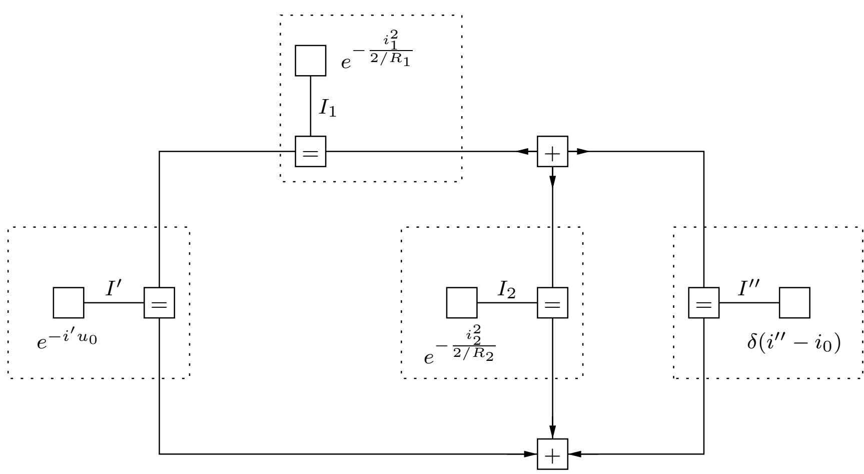


Figure 7: FFG (current version) corresponding to the electrical network in Fig. 4.

the branch currents) is normally found by writing up the Kirchhoff voltage law (KVL) equations, the Kirchhoff current law (KCL) equations, and the component law equations and finally solving them[2]. It is now possible to formulate two optimization problems (which are duals of each other), whose solution give either the voltages (in the first case) or the currents (in the second case) of the above electrical network. In the first case, one optimizes a function whose arguments are voltages which are subject only to the KVL equations (but not the KCL equations). In the second case, one optimizes a function whose arguments are currents which are subject only to the KCL equations (but not the KVL equations). In both cases, the function to be optimized is derived from the component laws. For details we refer to [8, 6, 3].

Fig. 4 shows an electrical network containing an ideal voltage source, an ideal current source, and two resistors. The factor graph that would correspond to the KVL equations, the KCL equations, and the component law equations is shown in Fig. 5: the global function is non-zero if the voltages and currents correspond to a solution of the KVL, KCL, and component law equations, otherwise it is zero.

Fig. 6 shows the FFG corresponding to the first optimization problem mentioned above. Here, the global function is a non-negative function and attains its maximum for the branch voltages that correspond to the branch voltages of the solution to the original electrical network problem. Similarly, Fig. 7 shows the FFG corresponding to the second optimization problem where the global function attains its maximum for the branch currents that correspond to the branch currents of the solution to the original electrical network problem. In the last two cases, the non-negative global function can be interpreted as (scaled) probability density functions of the occuring variables. More details about this example can be found in [3]. △

To conclude this section, we would like to point out additional topics.

- As already mentioned above, the two optimization problems are duals of each other [8, 3]. Additionally, 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. For more details, see e.g. [3, 6].

- One can derive an electrical network corresponding to the Kalman filtering problem, see e.g. [9, 6].

- The simplification of electrical network corresponds to performing the max-product algorithm. In the case of jointly Gaussian random variables, this is equivalent to the sum-product algorithm. Simplifying e.g. the electrical network which corresponds to the Kalman filtering problem results in the Kalman filtering algorithm. For more details, see e.g. [6, 5], but also Carter [10].

3 Connections between FFGs, Electrical Networks, and Entropies

In this section we will only be interested in jointly Gaussian random variables. First we review some results about differential entropies; then we will relate entropies (involving

[2]For the discusssion here we assume that the electrical network has a solution which is unique.

random variables appearing in an FFGs) to some effective resistances in an electrical network (where the electrical network was derived from that FFG).

Let $Z \sim \mathcal{N}(m, \sigma^2)$ have a Gaussian distribution with mean m and variance σ^2. The differential entropy is then (see e.g. Ch. 9 in [7])[3]

$$h(Z) = -\int_{-\infty}^{+\infty} p(z) \log p(z) \, \mathrm{d}z = \frac{1}{2} \log(2\pi e \sigma^2).$$

In general, if $\mathbf{Z} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$ has an n-dimensional multivariate Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{K}$, the (differential) entropy is $\frac{1}{2}\log((2\pi e)^n |\mathbf{K}|)$. A key ingredient will be the following fact (mentioned e.g. in Ch. 16.9 in [7]). If $Z_1, Z_2, \ldots, Z_n$ are jointly Gaussian distributed, then

$$h(Z_n | Z_1, Z_2, \ldots, Z_{n-1}) = \frac{1}{2} \log(2\pi e \hat{\sigma}_n^2),$$

where $\hat{\sigma}_n^2$ is the minimum mean squared error $E[(Z_n - \hat{Z}_n)^2]$ over all linear estimators $\hat{Z}_n$ based on $Z_1, Z_2, \ldots, Z_{n-1}$ (note that the linear estimator is optimal in the case of jointly Gaussian random variables). We even have the result that

$$h(Z_n | Z_1, Z_2, \ldots, Z_{n-1}) = h(Z_n | Z_1 = z_1, Z_2 = z_2, \ldots, Z_{n-1} = z_{n-1}) = \frac{1}{2} \log(2\pi e \hat{\sigma}_n^2),$$

which holds independently of the values of the measurements $Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1}$ as long as they do not contradict each other.

Combining these properties of Gaussian random variables with the result of Sec. 5.3 in [6] (about the relation of certain effective resistances and elements of the error covariance matrix)[4], we claim the following procedure for obtaining certain differential entropies from effective resistances.

Procedure 3 (Determination of $h(Z_n | Z_1, \ldots, Z_{n-1})$) *We make the following assumptions.*

- *We have an FFG (possibly with loops!) whose global function is proportional to a jointly Gaussian density.*
- *We are interested in the differential entropy $h(Z_n | Z_1, \ldots, Z_{n-1})$, where $Z_1, \ldots, Z_n$ are some of the branch labels in the FFG.*
- *An electrical network (containing linear resistors, ideal voltage sources, and DC transformers) has been derived from the FFG as was done in Ex. 1 ("voltage-mode derivation").*

To determine $h(Z_n | Z_1, \ldots, Z_{n-1})$, we can apply the following steps.

- *Replace all voltage sources by an open circuit.*

[3] The base of the logarithm is chosen to be 2.

[4] In this text, we use the term effective resistance; it is equivalent to the term input resistance as used in [6]. Note that by "measuring the effective resistance at a branch" we will mean that we measure the effective resistance between the two end nodes of that branch.

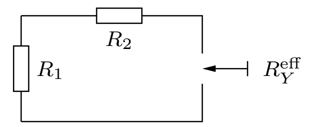
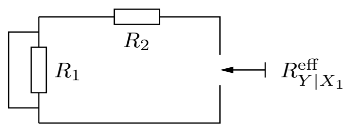
Electrical Network	Effective Resistance	Entropy
	$R_Y^{\text{eff}} = R_1 + R_2$	$h(Y) = \frac{1}{2}\log(2\pi e R_Y^{\text{eff}})$
	$R_{Y\|X_1}^{\text{eff}} = R_2$	$h(Y\|X_1) = \frac{1}{2}\log(2\pi e R_{Y\|X_1}^{\text{eff}})$

Table 1: Entropies $h(Y)$ and $h(Y|X_1)$. Here $R_1 = \sigma_1^2$ and $R_2 = \sigma_2^2$.

- *Measure the effective resistance $R^{\text{eff}}_{Z_n|Z_1\ldots Z_{n-1}}$ at the branch corresponding to Z_n, while short circuiting the elements corresponding to $Z_1, \ldots, Z_{n-1}$.*
- *The desired differential entropy is*

$$h(Z_n|Z_1, \ldots, Z_{n-1}) = \frac{1}{2}\log\left(2\pi e R^{\text{eff}}_{Z_n|Z_1\ldots Z_{n-1}}\right)$$

Example 4 [From effective resistances to differential entropies] We consider the FFG and its corresponding electrical network in Ex. 1. Applying Proc. 3, we start with the electrical network in Fig. 2 and replace the voltage source by an open circuit.

- The entropy $h(Y)$ is now related to the effective resistance R_Y^{eff} measured at the branch in the electrical network corresponding to Y (see Tab. 1).
- The entropy $h(Y|X_1)$ is related to the effective resistance $R^{\text{eff}}_{Y|X_1}$ measured at the same place but now with short circuiting the element corresponding to X_1 (see Tab. 1).
- Based on these two results, we can e.g. calculate the mutual information between X_1 and Y.

$$\begin{aligned} I(X_1;Y) = h(Y) - h(Y|X_1) &= \frac{1}{2}\log\left(\frac{2\pi e R_Y^{\text{eff}}}{2\pi e R^{\text{eff}}_{Y|X_1}}\right) \\ &= \frac{1}{2}\log\left(\frac{R_Y^{\text{eff}}}{R^{\text{eff}}_{Y|X_1}}\right) = \frac{1}{2}\log\left(\frac{\sigma_1^2+\sigma_2^2}{\sigma_2^2}\right). \end{aligned}$$

- From Rayleigh's Monotonicity Law[5] we must have $R_Y^{\text{eff}} \geq R^{\text{eff}}_{Y|X_1}$. Therefore $I(X_1;Y) \geq 0$, a well-know result from information theory (see e.g. [7]).

$\triangle$

[5] Rayleigh's Monotonicity Law (see e.g. [11]) states that if the resitances of a circuit are increased, the effective resistance between any two points can only increase. If they are decreased, it can only decrease.

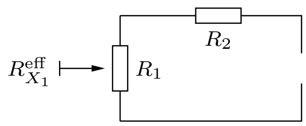
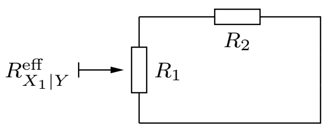
Electrical Network	Effective Resistance	Entropy
	$R^{\mathrm{eff}}_{X_1} = R_1$	$h(X_1) = \frac{1}{2}\log(2\pi e R^{\mathrm{eff}}_{X_1})$
	$R^{\mathrm{eff}}_{X_1\|Y} = (R_1\|\|R_2) = \frac{1}{\frac{1}{R_1}+\frac{1}{R_2}}$	$h(X_1\|Y) = \frac{1}{2}\log(2\pi e R^{\mathrm{eff}}_{X_1\|Y})$

Table 2: Entropies $h(X_1)$ and $h(X_1|Y)$. Here $R_1 = \sigma_1^2$ and $R_2 = \sigma_2^2$.

Example 5 [From effective resistances to differential entropies] We consider the FFG and its corresponding electrical network in Ex. 1. Applying Proc. 3, we start with the electrical network in Fig. 2 and replace the voltage source by an open circuit.

- The entropy $h(X_1)$ is now related to the effective resistance $R^{\mathrm{eff}}_{X_1}$ measured at the branch in the electrical network corresponding to X_1 (see Tab. 2).
- The entropy $h(X_1|Y)$ is related to the effective resistance $R^{\mathrm{eff}}_{X_1|Y}$ measured at the same place but now with short circuiting the element corresponding to Y (see Tab. 2).
- Based on these two results, we can e.g. calculate the mutual information between X_1 and Y.

$$I(X_1;Y) = h(X_1) - h(X_1|Y) = \frac{1}{2}\log\left(\frac{2\pi e R^{\mathrm{eff}}_{X_1}}{2\pi e R^{\mathrm{eff}}_{X_1|Y}}\right)$$
$$= \frac{1}{2}\log\left(\frac{R^{\mathrm{eff}}_{X_1}}{R^{\mathrm{eff}}_{X_1|Y}}\right) = \frac{1}{2}\log\left(\frac{\sigma_1^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right) = \frac{1}{2}\log\left(\frac{\sigma_1^2+\sigma_2^2}{\sigma_2^2}\right).$$

The final result is of course the same as in Ex. 4.

$\triangle$

Other relationships between results in electrical network theory and about entropies can be found. Let us mention two of them.

- (From electrical networks to entropies.) Shannon and Hagelbarger [12] gave some concavity result of effective resistances; their result can be stated as follows. We are interested in the effective resistance between two nodes in some given electrical network topology where we vary the resistances in the branches. Let R^{eff}_1 be the effective resistance for a first resistance configuration, and let R^{eff}_2 be the effective resistance for a second resistance configuration. Finally, let R^{eff} be the effective resistance where each resistance is the sum of the corresponding resistances in the first two configurations. Then,

$$R^{\text{eff}} \geq R_1^{\text{eff}} + R_2^{\text{eff}}.$$

(When slightly reformulating the setup, one gets a concavity result.) Let $h = (1/2)\log(2\pi e R^{\text{eff}})$, $h_1 = (1/2)\log(2\pi e R_1^{\text{eff}})$, and $h_2 = (1/2)\log(2\pi e R_2^{\text{eff}})$. Then

$$2^{2h} \geq 2^{2h_1} + 2^{2h_2},$$

which can be considered as sort of an extension of the entropy power inequality (see e.g. p.496 in [7]) in the jointly Gaussian case.

- (From entropies to electrical networks.) Entropy inequalities in Ch. 16 of [7] can be translated to network theory results.

We would like to conclude this section by pointing out that these entropy considerations were originally motivated and inspired by the results of Arnold and Loeliger [13] and Pfister et al. [14] who used the sum-product algorithm on a loopless FFG (representing a finite-state time-invariant system) to determine some entropies and information rates.

4 Conclusions

We hope that we could convince the reader that there are indeed connections between factor graphs, electrical networks, and entropies. As an open problem, we would like to mention that we wonder whether one can derive similar results as in Sec. 3 also in the non-Gaussian case; or if at least one can give lower or upper bounds on differential entropies and mutual informations.

Acknowlegments

We would like to thank Andi Loeliger for interesting and inspiring discussions.

References

[1] *Special Issue on "Codes on graphs and iterative decoding"*, vol. IT–47 of *IEEE Trans. on Inform. Theory*, Feb. 2001.

[2] 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.

[3] P. O. Vontobel and H.-A. Loeliger, "On factor graphs and electrical networks," in *Mathematical Systems Theory in Biology, Communication, Computation, and Finance, IMA Volumes in Math. & Appl.* (D. Gilliam and J. Rosenthal, eds.), Springer Verlag. To appear.

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

[5] 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.

[6] 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`.

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

[8] J. B. Dennis, *Mathematical Programming and Electrical Networks.* The Technology Press of The Massachusetts Institute of Technology, Cambridge, Mass., 1959.

[9] P. O. Vontobel, D. Lippuner, and H.-A. Loeliger, "Kalman filters, factor graphs, and electrical networks," in *Proc. Fifteenth Intern. Symp. on Math. Theory of Networks and Systems*, (Notre Dame, IN, USA), Aug. 12–16 2002. Available online under `http://www.nd.edu/~mtns/`.

[10] 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.

[11] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*, vol. 22 of *Carus Mathematical Monographs.* Washington, DC: Mathematical Association of America, 1984.

[12] C. E. Shannon and D. W. Hagelbarger, "Concavity of resistance functions," *Journal of Applied Physics*, vol. 27, pp. 42–43, 1956.

[13] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. 2001 IEEE Int. Conf. on Communications*, (Helsinki, Finland), pp. 2692–2695, June 11–14 2001.

[14] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," in *Proc. IEEE GLOBECOM*, (San Antonio, TX), pp. 2992–2996, Nov. 2001.