Factor Graph Duality

Lecture Notes for: Graphical Models and Inference
Henry D. Pfister

September 24th, 2014 (rev. 0)

1 Introduction

Many problems in the mathematical sciences can be simplified by a well-chosen change of variables and the marginalization problem for factor graphs is no different. In these notes, we outline the transformation process first for a single factor and then for the whole factor graph. The approach taken is based on analyzing the equations involved. A more intuitive graphical approach is taken in [1].

2 Transforming a Single Factor

2.1 Basic Idea

Let $\mathcal{X} = \{0, 1, \ldots, q-1\}$ be a finite alphabet and consider the function $f : \mathcal{X}^n \to \mathbb{R}$. Now, suppose that we want to compute the sum

$$Z(\underline{\mu}) \triangleq \sum_{x_1^n \in \mathcal{X}^n} f(x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} \mu_j(x_j)$$

for the vector $\underline{\mu} = (\mu_1, \ldots, \mu_n)$ where each element is a function $\mu_j : \mathcal{X} \to \mathbb{R}$. Using an invertible $q \times q$ matrix A, the same quantity can be written as

$$Z(\underline{\mu}) = \sum_{x_1^n \in \mathcal{X}^n} f(x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} \sum_{\hat{x} \in \mathcal{X}} A_{x_j, \hat{x}} \sum_{x \in \mathcal{X}} A^{-1}_{\hat{x}, x} \mu_j(x).$$

If we define $\underline{\hat{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)$ by $\hat{\mu}_j(\hat{x}_j) \triangleq \sum_{x \in \mathcal{X}} A^{-1}_{\hat{x}, x} \mu_j(x)$, then we can rewrite this as

$$\begin{aligned} Z(\underline{\mu}) &= \sum_{x_1^n \in \mathcal{X}^n} f(x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} \sum_{\hat{x}_j \in \mathcal{X}} A_{x_j, \hat{x}_j} \hat{\mu}_j(\hat{x}_j) \\ &= \sum_{\hat{x}_1^n \in \mathcal{X}^n} \underbrace{\left[\sum_{x_1^n \in \mathcal{X}^n} f(x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} A_{x_j, \hat{x}_j} \right]}_{\triangleq \hat{f}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)} \prod_{j=1}^{n} \hat{\mu}_j(\hat{x}_j) \\ &= \sum_{\hat{x}_1^n \in \mathcal{X}^n} \hat{f}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \prod_{j=1}^{n} \hat{\mu}_j(\hat{x}_j) \triangleq \hat{Z}(\underline{\hat{\mu}}), \end{aligned}$$

where $\hat{f}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$ is the transformed factor. To see this in the context of marginalization, we define

$$\underline{\mu}\Big|_{\mu_k(x_k) = \chi(x_k, x_k')} (\underline{x}) \triangleq (\mu_1(x_1), \ldots, \mu_{k-1}(x_{k-1}), \chi(x_k, x_k'), \mu_{k+1}(x_{k+1}), \ldots, \mu_n(x_n))$$

and observe that

$$\begin{aligned} g_k(x'_k) &\triangleq \sum_{x_1^n \in \mathcal{X}^n} \delta_{x'_k,x_k} f(x_1,x_2,\ldots,x_n) \prod_{j=1,j\neq k}^{n} \mu_j(x_j) \\ &= Z\Big(\underline{\mu}\Big|_{\mu_k(x_k)=\delta_{x_k,x'_k}}\Big) = \hat{Z}\Big(\underline{\hat{\mu}}\Big|_{\hat{\mu}_k(\hat{x}_k)=A^{-1}_{\hat{x}_k,x'_k}}\Big) \\ &= \sum_{\hat{x}_1^n \in \mathcal{X}^n} A^{-1}_{\hat{x}_k,x'_k} \hat{f}(\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_n) \prod_{j=1,j\neq k}^{n} \hat{\mu}_j(\hat{x}_j) \\ &= \sum_{\hat{x}'_k \in \mathcal{X}} A^{-1}_{\hat{x}'_k,x'_k} \underbrace{\sum_{\hat{x}_1^n \in \mathcal{X}^n} \delta_{\hat{x}'_k,\hat{x}_k} \hat{f}(\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_n) \prod_{j=1,j\neq k}^{n} \hat{\mu}_j(\hat{x}_j)}_{\triangleq \hat{g}_k(\hat{x}'_k)}. \end{aligned}$$

Notice that $g_k(x'_k)$ is the belief-propagation (BP) message from f to x_k for the original factor when the input messages are $\mu_j(x_j)$. Likewise, $\hat{g}_k(\hat{x}'_k)$ gives the BP message for $\hat{f}$ to x_k for the transformed factor when the input messages are $\hat{\mu}_j(\hat{x}_j) = \sum_{x\in\mathcal{X}} A^{-1}_{\hat{x},x}\mu_j(x)$. Thus, this transformation constitutes a change of basis for the BP messages.

2.2 Connection to Duality

This technique is quite useful $\mathcal{X}$ is a finite field and $f(x_1,\ldots,x_n)$ is the indicator function of a subspace $S \subseteq \mathcal{X}^n$. In this case, the matrix A is typically chosen to be the Fourier transform associated with the additive group of $\mathcal{X}$. With this choice, $\hat{f}$ is called the dual factor of f and $\hat{f}(\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_n)$ becomes a scaled indicator function for the dual space $S^{\perp}$.

Consider the finite field with $|\mathcal{X}| = q = p^m$ elements for prime p. It is well-known that the additive group of $\mathcal{X}$ is isomorphic to the set $\{0,1,\ldots,p-1\}^m$ of vectors with elementwise modulo-p addition. Thus, we assume wolog that $\mathcal{X} = \{0,1,\ldots,p-1\}^m$ and define the Fourier transform

$$A_{x,\hat{x}} = \frac{1}{\sqrt{q}} e^{-2\pi i \langle x,\hat{x}\rangle/p},$$

where $\langle x,\hat{x}\rangle_{\mathcal{X}}$ is the standard inner product between these two length-m vectors. Using this convention,

$$\sum_{\hat{x}\in\mathcal{X}} A_{x,\hat{x}} A^{-1}_{\hat{x},x'} = \delta_{x,x'}.$$

To see the duality between indicator functions, we let the subspace $S = \big\{uG \,|\, u \in \mathcal{X}^k\big\}$ be defined by a $k\times n$ generator matrix G over $\mathcal{X}$ and we extend the inner product to $\mathcal{X}^n$ with $\langle \underline{x}, \underline{\hat{x}}\rangle_{\mathcal{X}^n} \triangleq \sum_{j=1}^{n} \langle x_j, \hat{x}_j\rangle_{\mathcal{X}}$. Then, we can write

$$\begin{aligned} \hat{f}(\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_n) &= \sum_{x_1^n \in \mathcal{X}^n} f(x_1,x_2,\ldots,x_n) \prod_{j=1}^{n} A_{x_j,\hat{x}_j} \\ &= \sum_{x_1^n \in \mathcal{X}^n} f(x_1,x_2,\ldots,x_n) \prod_{j=1}^{n} \frac{1}{\sqrt{q}} e^{-2\pi i \langle x_j,\hat{x}_j\rangle_{\mathcal{X}}/p} \\ &= q^{-n/2} \sum_{u_1^n \in \mathcal{X}^k} e^{-\frac{2\pi i}{p}\sum_{j=1}^{n}\langle x_j,\hat{x}_j\rangle_{\mathcal{X}}} \\ &= q^{-n/2} \sum_{u_1^n \in \mathcal{X}^k} e^{-\frac{2\pi i}{p}\langle \underline{u}G,\underline{\hat{x}}\rangle_{\mathcal{X}^n}} \\ &= \begin{cases} q^{-n/2}q^k & \text{if } \langle \underline{u}G,\underline{\hat{x}}\rangle_{\mathcal{X}^n} = 0 \text{ for all } \underline{u} \in \mathcal{X}^k \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The first case holds because, if $\langle \underline{u}G,\underline{\hat{x}}\rangle_{\mathcal{X}^n} = 0$ for all $\underline{u}$, then the exponential is 1 for all $\underline{u}$ and the sum has q^k terms. For the second case, we observe that, if there is some $\underline{u}$ such that $\langle \underline{u}G,\underline{\hat{x}}\rangle_{\mathcal{X}^n} \neq 0$, then

$$\big\{u_1^n \in \mathcal{X}^k \mid \langle \underline{u}G,\underline{\hat{x}}\rangle_{\mathcal{X}^n} = a\big\} = a \cdot \big\{u_1^n \in \mathcal{X}^k \mid \langle \underline{u}G,\underline{\hat{x}}\rangle_{\mathcal{X}^n} = 1\big\}$$

for all $a \in \mathcal{X}$. Thus, each set has the same size (e.g., q^{k-1}) and we get

$$\begin{aligned}\sum_{u_1^n \in \mathcal{X}^k} e^{-\frac{2\pi i}{p} \langle \underline{u}G, \underline{\hat{x}} \rangle_{\mathcal{X}^n}} &= \sum_{a \in \mathcal{X}} \sum_{u_1^k : \langle \underline{u}G, \underline{\hat{x}} \rangle = a} e^{-\frac{2\pi i}{p} \langle \underline{u}G, \underline{\hat{x}} \rangle_{\mathcal{X}^n}} \\ &= \sum_{a \in \mathcal{X}} \sum_{u_1^k : \langle \underline{u}G, \underline{\hat{x}} \rangle_{\mathcal{X}^n} = a} e^{-\frac{2\pi i}{p} a} \\ &= \sum_{a \in \mathcal{X}} q^{k-1} e^{-\frac{2\pi i}{p} a} \\ &= 0.\end{aligned}$$

Since the dual code is defined to be the set of all vectors whose inner product with all codewords is 0, we see that $\hat{f}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$ is $q^{k-n/2}$ times the indicator function for the dual code $\mathcal{C}^{\perp}$.

2.3 Applications

The results in the previous section have two immediate applications. First, they allow one to write Symbol-APP decoding problem for a linear code in terms of the Symbol-APP decoding problem for the dual code. This reduces the complexity of brute-force decoding from q^k to q^{n-k}. Second, this result provides a proof of the famous MacWilliams identity relating the weight enumerator of a linear code to the weight enumerator of its dual code.

2.3.1 Decoding

For the first application, let X_1^n be a random vector whose distribution is defined by $\Pr(X_1^n = x_1^n) \propto f(x_1, \ldots, x_n)$ and let Y_1^n be an observation of X_1^n through a discrete memoryless channel with transition probabilities $W(y|x) = \Pr(Y = y|X = x)$. Then, the choice $\mu_j(x_j) = W(Y_j|x_j)$ implies that

$$\Pr\left(X_k = x_k' \,\middle|\, Y_1^n\right) \propto \mu_k(x_k') g_k(x_k').$$

Transforming f with the Fourier transform implies that

$$g_k(x_k') = \sum_{\hat{x}_k' \in \mathcal{X}} A^{-1}_{\hat{x}_k', x_k'} \hat{g}_k(\hat{x}_k') = \sum_{\hat{x}_k' \in \mathcal{X}} \frac{1}{\sqrt{q}} e^{2\pi i \left\langle x_k', \hat{x}' \right\rangle / p} \hat{g}_k(\hat{x}_k').$$

Thus, the marginalization of $\hat{f}$ with the messages

$$\mu_j(\hat{x}_j) = \sum_{x \in \mathcal{X}} \frac{1}{\sqrt{q}} e^{2\pi i \langle x, \hat{x}_j \rangle / p} \mu_j(x)$$

allows one to compute the marginalization of f for the messages $\mu_j(x_j)$. Putting these equations together gives

$$\mu_k(x_k') = \sum_{\hat{x}_k' \in \mathcal{X}} \frac{1}{\sqrt{q}} e^{2\pi i \left\langle x_k', \hat{x}' \right\rangle / p} \sum_{\hat{x}_1^n \in \mathcal{X}^n} \delta_{\hat{x}_k', \hat{x}_k} \hat{f}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \prod_{j=1, j \neq k}^{n} \left(\sum_{x \in \mathcal{X}} \frac{1}{\sqrt{q}} e^{2\pi i \langle x, \hat{x}_j \rangle / p} \mu_j(x) \right).$$

This approach was introduced for linear codes in 1976 by Hartmann and Rudolph [2].

As you will see in the next example, if $f(x_1, \ldots, x_n)$ is the indicator function of a linear code, then $\hat{f}(x_1, \ldots, x_n)$ is proportional to the indicator function of the dual code. Thus, if we let $f(x_1, \ldots, x_n) = I(x_1 \oplus \cdots \oplus x_n)$ be the even-parity factor, then $\hat{f}(x_1, \ldots, x_n) \propto I(x_1 = \cdots = x_n)$ because the repeat code is dual to the even-parity code. In this case, the previous equation simplifies to

$$\mu_k(x_k') = \sum_{\hat{x}_k' \in \mathcal{X}} \frac{1}{\sqrt{q}} e^{2\pi i \left\langle x_k', \hat{x}' \right\rangle / p} \prod_{j=1, j \neq k}^{n} \left(\sum_{x \in \mathcal{X}} \frac{1}{\sqrt{q}} e^{2\pi i \left\langle x, \hat{x}_k' \right\rangle / p} \mu_j(x) \right).$$

For the binary case, this implies that

$$\mu_k(x_k') \propto \left[\prod_{j=1, j \neq k}^{n} \left(\mu_j(0) + \mu_j(1)\right) + (-1)^{x_k'} \prod_{j=1, j \neq k}^{n} \left(\mu_j(0) - \mu_j(1)\right) \right].$$

2.3.2 The MacWilliams Identity

For the second application, let the weight enumerator of a linear code be $A(z) = \sum_{h=0}^{n} A_h z^h$ where A_h is the number of codewords with h non-zero positions. Similarly, the weight enumerator of the dual code be $A^{\perp}(z) = \sum_{h=0}^{n} A_h^{\perp} z^h$ where $A_h^{\perp}$ is the number of codewords in the dual code with h non-zero positions. For the binary case, MacWilliams identity $A(z) = (1+z)^n 2^{k-n} A^{\perp}(\frac{1+z}{1-z})$ follows from letting f be the indicator function of code and writing

$$\begin{aligned}
A(z) &= \sum_{h=0}^{n} A_h z^h \\
&= \sum_{x_1^n \in \mathcal{X}^n} f(x_1, x_2, \ldots, x_n) \prod_{j=1}^{n} z^{x_j} \\
&= \sum_{\hat{x}_1^n \in \mathcal{X}^n} \hat{f}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \prod_{j=1}^{n} \frac{1}{\sqrt{2}}(1+z) \left(\frac{1-z}{1+z}\right)^{\hat{x}_j} \\
&= \sum_{\hat{x}_1^n \in \mathcal{X}^n} 2^{k-n/2} I\left(\hat{x}_1^n \in \mathcal{C}^{\perp}\right) \prod_{j=1}^{n} \frac{1}{\sqrt{2}}(1+z) \left(\frac{1-z}{1+z}\right)^{\hat{x}_j} \\
&= (1+z)^n 2^{k-n} \sum_{h=0}^{n} A_h^{\perp} \left(\frac{1-z}{1+z}\right)^h \\
&= (1+z)^n 2^{k-n} A^{\perp}\left(\tfrac{1+z}{1-z}\right).
\end{aligned}$$

For the third step, we use the fact that $\mu_j(x_j) = z^{x_j}$ has the Fourier transform

$$\hat{\mu}_j(\hat{x}_j) = \begin{cases} \frac{1}{\sqrt{2}}(1+z) & \text{if } \hat{x}_j = 0 \\ \frac{1}{\sqrt{2}}(1-z) & \text{if } \hat{x}_j = 1, \end{cases}$$

which can be written compactly as $\hat{\mu}_j(\hat{x}_j) = (1+z)\left((1-z)/(1+z)\right)^{\hat{x}_j}$. This identity was introduced by MacWilliams in [3].

3 Transforming the Whole Factor Graph

One can apply similar techniques to the whole factor graph. In that case, it is useful to consider the special case of normal factor graphs. For details, see [4, 1].

References

[1] G. D. F. Jr. and P. O. Vontobel, "Partition functions of normal factor graphs," *arXiv preprint arXiv:1102.0316*, 2011.

[2] C. R. P. Hartmann and L. D. Rudolph, "An optimum symbol-by-symbol decoding rule for linear codes," *IEEE Trans. Inform. Theory*, vol. 22, no. 5, pp. 514–517, 1976.

[3] J. MacWilliams, "A theorem on the distribution of weights in a systematic code," *The Bell Syst. Techn. J.*, vol. 42, no. 1, pp. 79–94, 1963.

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