Information-entropy trade-off for random vectors
From: C. M. Caves
To: T. Brun, A. J. Scott, and R. Schack
2005 September 14; minor updates 2006 March 11

The results here are based on formulae in the two appendices in R. Schack, C. M. Caves, and G. M. D'Ariano, "Hypersensitivity to perburbation in the quantum kicked top," Phys. Rev. E **50**, 972–987 (1994). The relevant equations in the two appendices are (A18) and (B5)–(B6). A preliminary version of the analysis here can be found at the end of Sec. V of R. Schack and C. M. Caves, "Information-theoretic characterization of quantum chaos," Phys. Rev. E **53**, 3257–3270 (1996). This analysis was further elaborated to essentially the present form in A. N. Soklakov and R. Schack, "Preparation information and optimal decompositions for mixed quantum states," J. Mod. Opt. **47**, 2265–2276 (2000).

Consider N vectors randomly distributed in a D-dimensional Hilbert space, where we assume that $N \geq D$. Given an entropy $H \leq \log D$, we want to group the vectors into groups that on average have this entropy and then ask how much information I is required to specify a group. The relation between I and H is the *information-entropy trade-off.*

We can estimate this trade-off by considering a grouping of the vectors into spheres on projective Hilbert-space whose radius is given by a Hilbert-space angle ϕ. The number of spheres of radius ϕ is [Eq. (A18) of Schack1994]

$$N_D(\phi) = \frac{\mathcal{V}_D}{\mathcal{V}_D(\phi)} = (\sin^2\phi)^{-(D-1)} , \tag{1}$$

where $\mathcal{V}_D(\phi)$ is the volume of a sphere of radius ϕ and $\mathcal{V}_D$ is the total volume of projective Hilbert space. The entropy of vectors distributed randomly within a sphere of radius ϕ is [Eqs. (B5)–(B6) of Schack1994]

$$H_D(\phi) = -\lambda_0 \log \lambda_0 - (1-\lambda_0) \log\left(\frac{1-\lambda_0}{D-1}\right) = H_2(\lambda_0) + (1-\lambda_0)\log(D-1) , \tag{2}$$

where $H_2(\lambda_0)$ is the binary entropy corresponding to the largest eigenvalue

$$\lambda_0 = 1 - \frac{D-1}{D}\sin^2\phi \qquad \Longrightarrow \qquad \frac{1-\lambda_0}{D-1} = \frac{\sin^2\phi}{D} \leq \frac{1}{D} . \tag{3}$$

For a long time, I thought we could approximate (2) as

$$H_D(\phi) \sim \sin^2\phi \log D , \tag{4}$$

but a little bit of plotting in Mathematica shows this gives a hopelessly bad approximation, especially near $\phi = \pi/2$, which is the most important place. Here we will use the exact expression for $H_D(\phi)$.

Now we're ready to get started. We group the N vectors into groups of radius ϕ. The number of vectors per group is

$$N_V(\phi) = \frac{N}{N_D(\phi)} = N(\sin^2\phi)^{D-1} , \tag{5}$$

provided this number is not less than one. There is clearly a critical angle, ϕ_b, at which there is only one vector per group, i.e.,

$$N_V(\phi_b) = 1 \qquad \Longleftrightarrow \qquad N_D(\phi_b) = N \qquad \Longleftrightarrow \qquad (\sin^2\phi_b)^{D-1} = \frac{1}{N} \; . \tag{6}$$

For $\phi \geq \phi_b$, there are $N_D(\phi)$ groups, each containing approximately $N_V(\phi)$ vectors, but for $\phi \leq \phi_b$, there are N groups, each containing one vector.

The information required to specify a group at resolution angle ϕ is thus

$$I(\phi) = \begin{cases} \log N \; , & \phi \leq \phi_b, \\ \log N_D(\phi) = -(D-1)\log(\sin^2\phi) \; , & \phi \geq \phi_b. \end{cases} \tag{7}$$

There is clearly another critical angle, ϕ_d, at which there are only two groups, i.e.,

$$N_V(\phi_d) = N/2 \quad \Longleftrightarrow \quad N_D(\phi_d) = 2 \quad \Longleftrightarrow \quad (\sin^2\phi_d)^{D-1} = \frac{1}{2} \quad \Longleftrightarrow \quad I(\phi_d) = 1 \; . \tag{8}$$

For $\phi \geq \phi_d$, we can't talk about grouping the vectors into spheres of equal radius, so we aren't justified in saying that $I(\phi)$ is given by $\log N_D(\phi)$. To deal with this, we give up on describing the situation $\phi > \phi_d$, amending Eq. (7) to be

$$I(\phi) = \begin{cases} \log N \; , & \phi \leq \phi_b, \\ \log N_D(\phi) = -(D-1)\log(\sin^2\phi) \; , & \phi_b \leq \phi \leq \phi_d. \end{cases} \tag{9}$$

Notice that for $\phi_b \leq \phi \leq \phi_d$, we have

$$\sin^2\phi = 2^{-I/(D-1)} = e^{-I\ln 2/(D-1)} \qquad \Longrightarrow \qquad \begin{array}{l} \lambda_0 = 1 - \dfrac{D-1}{D} 2^{-I/(D-1)} \\ \dfrac{1-\lambda_0}{D-1} = \dfrac{\sin^2\phi}{D} = \dfrac{2^{-I/(D-1)}}{D} \end{array} . \tag{10}$$

This gives

$$\sin^2\phi_d = 2^{-1/(D-1)} \; , \tag{11}$$

so that as long as $D \gg 1$,

$$\sin^2\phi_d \simeq 1 - \frac{\ln 2}{D-1} \qquad \Longrightarrow \qquad \phi_d \simeq \frac{\pi}{2} - \sqrt{\frac{\ln 2}{D-1}} \; . \tag{12}$$

Equation (10) shows that there are two important cases in terms of the number of vectors. If $\log N \ll D$ ($N \ll 2^D$), a situation we refer to as *sparse* collection of random vectors (even though there can be a heck of a lot of them), we have $I \leq \log N \ll D$, giving

$$\sin^2\phi \simeq 1 - \frac{I\ln 2}{D-1} \qquad \Longrightarrow \qquad \phi \simeq \frac{\pi}{2} - \sqrt{\frac{I\ln 2}{D-1}} \tag{13}$$

over the entire range $\phi_b \le \phi \le \phi_d$. In particular, we have

$$\sin^2\phi_b \simeq 1 - \frac{\log N \ln 2}{D-1} = 1 - \frac{\ln N}{D-1} \qquad \Longrightarrow \qquad \phi_b \simeq \frac{\pi}{2} - \sqrt{\frac{\ln N}{D-1}} \ . \tag{14}$$

The number of groups increases so fast as ϕ retreats from $\pi/2$ that for a sparse collection, there is a group for each vector when the radius ϕ is still quite close to $\pi/2$ (Hilbert space is a big place!). In contrast, if $\log N \gg D$ ($N \gg 2^D$), which we call a *dense* collection of vectors, then

$$\phi_b \simeq \sin\phi_b = 2^{-\log N/2(D-1)} \ll 1 \ , \tag{15}$$

meaning that to get to one vector per group, the group radius ϕ_b must be small.

When we start thinking about the entropy of the groups, it becomes clear that there is yet another critical angle, ϕ_c, the angle at which the number of vectors per group equals the Hilbert-space dimension:

$$N_V(\phi_c) = D \quad \Longleftrightarrow \quad N_D(\phi_c) = \frac{N}{D} \quad \Longleftrightarrow \quad (\sin^2\phi_c)^{D-1} = \frac{D}{N} \quad \Longleftrightarrow \quad I(\phi_c) = \log N - \log D \ . \tag{16}$$

For $\phi \ge \phi_c$, there are sufficiently many vectors in each group to explore all the available dimensions, so the entropy will be close to the entropy $H_D(\phi)$ for a group of random vectors of radius ϕ in D dimensions. In contrast, for $\phi_b \le \phi \le \phi_c$, the vectors in a group can explore roughly only $N_V(\phi)$ dimensions, thus giving an entropy close to $H_{N_V(\phi)}(\phi)$. Finally, for $\phi \le \phi_b$, there is only one vector per group, so $H = 0$. Summarizing, we have

$$H(\phi) \simeq \begin{cases} 0 \ , & \phi \le \phi_b, \\ H_{N_V(\phi)}(\phi) \ , & \phi_b \le \phi \le \phi_c, \\ H_D(\phi) \ , & \phi \ge \phi_c. \end{cases} \tag{17}$$

Our main interest is in the relation between H and I, so we eliminate the resolution angle ϕ from the above expressions. The region $\phi \le \phi_b$ simply gives $H = 0$ and $I = \log N$. For $\phi_b \le \phi \le \phi_c$, i.e., $\log N \ge I \ge \log N - \log D$, we have

$$H \simeq H_{N_V(\phi)}(\phi) = H_2(\lambda) + (1-\lambda)\log\bigl(2^{-I}N - 1\bigr) \ , \qquad \begin{aligned} \lambda &= 1 - \frac{2^{-I}N - 1}{2^{-I}N} 2^{-I/(D-1)} \\ &= 1 - 2^{-I/(D-1)}\left(1 - \frac{2^I}{N}\right) \end{aligned} \ . \tag{18}$$

and for $\phi_c \le \phi \le \phi_d$, i.e., $\log N - \log D \ge I \ge 1$, we get

$$H \simeq H_D(\phi) = H_2(\lambda_0) + (1-\lambda_0)\log(D-1) \ , \quad \lambda_0 = 1 - \frac{D-1}{D} 2^{-I/(D-1)} \ . \tag{19}$$

Again summarizing, we have

$$H \simeq \begin{cases} H_2(\lambda) + (1-\lambda)\log\bigl(2^{-I}N - 1\bigr) \ , & \log N \ge I \ge \log N - \log D, \\ H_2(\lambda_0) + (1-\lambda_0)\log(D-1) \ , & \log N - \log D \ge I \ge 1. \end{cases} \tag{20}$$

with λ and λ_0 given by Eqs. (18) and (19). Equation (20) is the trade-off relation we are seeking.

The important part of the trade-off relation is the part that is independent of the number of random vectors, i.e., for $1 \le I \le \log N - \log D$. Notice that to investigate this region, we need $N \gg D$, but we do *not* need N so large that the perturbation samples typical vectors, i.e., $N \gtrsim 2^{2(D-1)}$, which would be a dense set of vectors. *We do not need a dense collection of vectors to investigate the important part of the trade-off relation.*

Before going further, let's put the trade-off relation (20) in other forms. First note that for the second case, which is the case of interest, we have

$$\begin{aligned} H_D(\phi) &= -\lambda_0 \log \lambda_0 - (1-\lambda_0)\log\left(\frac{1-\lambda_0}{D-1}\right) \\ &= -\lambda_0 \log \lambda_0 - (1-\lambda_0)\log\left(\frac{2^{-I/(D-1)}}{D}\right) \\ &= -\lambda_0 \log \lambda_0 + (1-\lambda_0)\left(\frac{I}{D-1} + \log D\right) \\ &= \log D - \lambda_0 \log(D\lambda_0) + \frac{I 2^{-I/(D-1)}}{D} \\ &= \log D - \frac{1}{D}\Big(\big(D(1-2^{-I/(D-1)}) + 2^{-I/(D-1)}\big) \\ &\qquad \times \log\big(D(1-2^{-I/(D-1)}) + 2^{-I/(D-1)}\big) - I 2^{-I/(D-1)}\Big)\ . \end{aligned} \tag{21}$$

For a sparse collection of vectors, for which $I \le \log N - \log D \le \log N \ll D$, or anytime we have $I \ll D$, we can approximate this by

$$H_D(\phi) \simeq \log D - \frac{1}{D}\big((1 + I\ln 2)\log(1 + I\ln 2) - I\big)\ . \tag{22}$$

This is a pretty good approximation for sparse collections, probably more than good enough given that the entire approach is only an approximation to an optimal grouping method for random vectors.

We can manipulate the first case in Eq. (20) in a similar way:

$$\begin{aligned} H_{N_V(\phi)}(\phi) &= -\lambda\log\lambda - (1-\lambda)\log\left(\frac{1-\lambda}{1-2^I/N}\right) + (1-\lambda)\log\left(\frac{N}{2^I}\right) \\ &= \log N - I - \lambda\log\left(\frac{N\lambda}{2^I}\right) - (1-\lambda)\log\left(\frac{1-\lambda}{1-2^I/N}\right)\ . \end{aligned} \tag{23}$$

The factor $2^I/N$ increases from $1/D$ for $I = \log N - \log D$ to 1 for $I = \log N$. For a sparse collection, i.e., $I \le \log N \ll D$, we can approximate λ by

$$\lambda \simeq \frac{2^I}{N} + \frac{I\ln 2}{D-1}\left(1 - \frac{2^I}{N}\right)\ . \tag{24}$$

The second term is always small. When the first term dominates, the second two terms in Eq. (23) vanish. When the first term is as small or smaller than the second, it is easy to see that the second two terms in Eq. (23) are both small. Thus for a sparse collection, it is always a good approximation to use

$$H_{N_V(\phi)}(\phi) = \log N - I \; . \tag{25}$$

The conclusion of these considerations is that for sparse collections, the trade-off relation (20) is well approximated by

$$H \simeq \begin{cases} \log N - I \; , & \log N \geq I \geq \log N - \log D, \\ \log D - \dfrac{1}{D}\big((1 + I \ln 2) \log(1 + I \ln 2) - I\big) \; , & \log N - \log D \geq I \geq 1. \end{cases} \tag{26}$$

The place where these approximate expressions are worst is at the knee between the two behaviors, which is also where the approximate treatment of the grouping is at its worst, so it doesn't matter much.

There are two good ways to capture the trade-off between information and entropy in a single number. The first is in terms of the derivative $(dI/dH)|_{I=1}$, which tells us the instantaneous information-entropy trade-off at the last place where our approximate expressions are valid. We can calculate the derivative from the following considerations. In the region of interest, which is the second case in Eq. (20), we have

$$\begin{aligned} \frac{dI(\phi)}{d\phi} &= \frac{-2(D-1)\sin\phi\cos\phi}{\sin^2\!\phi \ln 2} \; , \\ \frac{dH(\phi)}{d\phi} &= \left(\frac{dH_2(\lambda_0)}{d\lambda_0} - \log(D-1)\right)\frac{d\lambda_0}{d\phi} \\ &= \left[\log\left(\frac{1-\lambda_0}{\lambda_0}\right) - \log(D-1)\right]\left(-\frac{D-1}{D} 2\sin\phi\cos\phi\right) \\ &= 2\frac{D-1}{D}\sin\phi\cos\phi\log\left(\frac{\lambda_0}{(1-\lambda_0)/(D-1)}\right) \\ &= 2\frac{D-1}{D}\sin\phi\cos\phi\log\left(\frac{D}{\sin^2\!\phi} - (D-1)\right) \; . \end{aligned} \tag{27}$$

This gives us

$$\frac{dI}{dH} = -\frac{D/\sin^2\!\phi}{\ln\big(D/\sin^2\!\phi - (D-1)\big)} = -\frac{D2^{I/(D-1)}}{\ln\big(1 + D(2^{I/(D-1)} - 1)\big)} \; , \quad \log N - \log D \geq I \geq 1. \tag{28}$$

Notice that this derivative diverges at $I = 0$.

For a sparse collection of vectors, for which $I \leq \log N \ll D$, or anytime we have $I \ll D$, we can approximate Eq. (28) by

$$\frac{dI}{dH} \simeq -\frac{D}{\ln\left(1 + \dfrac{D}{D-1} I \ln 2\right)} \simeq -\frac{D}{\ln(1 + I\ln 2)} \; , \tag{29}$$

where the second approximation expression assumes $D \gg 1$. Notice that this second expression follows from differentiating Eq. (22) and retains the singularity at $I = 0$. The slope introduced above is thus

$$\left.\frac{dI}{dH}\right|_{I=1} = -\frac{D2^{1/(D-1)}}{\ln\bigl(1 + D(2^{1/(D-1)} - 1)\bigr)} \simeq -\frac{D}{\ln\left(1 + \dfrac{D}{D-1}\ln 2\right)} \simeq -\frac{D}{\ln(1+\ln 2)} = -1.899D\ . \tag{30}$$

The final approximate value on the right, which assumes $D \gg 1$, is, in fact, always off by less than 10% and is off by less than 1% for $D \geq 36$. Notice that for $I \ll 1$, we have

$$\frac{dI}{dH} = \frac{D-1}{I\ln 2}\ , \tag{31}$$

which shows the divergence in slope at $I = 0$.

The second good way to characterize the trade-off is in terms of a finite trade-off between information and entropy, $(\Delta I/\Delta H)|_{I=1}$, where

$$\frac{\Delta I}{\Delta H} = \frac{I}{H - \log D}\ . \tag{32}$$

To get this into a reasonable form, we use Eq. (21) to write

$$\frac{\Delta I}{\Delta H} = \frac{D}{2^{-I/(D-1)} - \dfrac{\bigl(D - (D-1)2^{-I/(D-1)}\bigr)\log\bigl(D - (D-1)2^{-I/(D-1)}\bigr)}{I}}\ . \tag{33}$$

For a sparse collection of vectors, for which $I \leq \log N \ll D$, or anytime we have $I \ll D$, we can approximate Eq. (33) by

$$\frac{\Delta I}{\Delta H} \simeq -\frac{D}{\dfrac{(1 + I\ln 2)\log(1 + I\ln 2)}{I} - 1}\ . \tag{34}$$

The trade-off introduced above is thus

$$\begin{aligned}\left.\frac{\Delta I}{\Delta H}\right|_{I=1} &= \frac{D}{2^{-1/(D-1)} - \bigl(D - (D-1)2^{-1/(D-1)}\bigr)\log\bigl(D - (D-1)2^{-1/(D-1)}\bigr)} \\ &\simeq -\frac{D}{(1+\ln 2)\log(1+\ln 2) - 1} = -3.493D\ .\end{aligned} \tag{35}$$

The approximate expression is off by less than 10% of the actual value for $D \geq 7$ and by less than 1% for $D \geq 58$. The ratio of the two measures, $(dI/dH)|_{I=1}/(\Delta I/\Delta H)_{I=1}$ decreases monotonically from a value of 0.687128 at $D = 2$ to an asymptotic value of

$$\frac{(1+\ln 2)\log(1+\ln 2) - 1}{\ln(1+\ln 2)} = 0.543681 \tag{36}$$

as D goes to infinity.

One has to be careful in using the approximation (34) for $I \ll 1$ because one needs to keep second-order terms in I in this situation. In particular, for $I \ll 1$, we have

$$\begin{aligned} D-(D-1)2^{-I/(D-1)} &\simeq D-(D-1)\left(1-\frac{I\ln 2}{D-1}+\frac{1}{2}\left(\frac{I\ln 2}{D-1}\right)^2\right) \\ &= 1+I\ln 2-\frac{1}{2}\frac{(I\ln 2)^2}{D-1}\;, \end{aligned} \tag{37}$$

which gives

$$\begin{aligned} \log\bigl(D-(D-1)2^{-I/(D-1)}\bigr) &\simeq \frac{1}{\ln 2}\left(I\ln 2-\frac{1}{2}\frac{(I\ln 2)^2}{D-1}-\frac{1}{2}(I\ln 2)^2\right) \\ &= I-\frac{1}{2}\frac{D}{D-1}I^2\ln 2\;. \end{aligned} \tag{38}$$

Plugging this into Eq. (33), we get for $I \ll 1$,

$$\begin{aligned} \frac{\Delta I}{\Delta H} &\simeq \frac{D}{1-\dfrac{I\ln 2}{D-1}-(1+I\ln 2)\left(1-\dfrac{1}{2}\dfrac{D}{D-1}I\ln 2\right)} \\ &\simeq \frac{D}{1-\dfrac{I\ln 2}{D-1}-1-I\ln 2+\dfrac{1}{2}\dfrac{D}{D-1}I\ln 2} \\ &= -\frac{2(D-1)}{I\ln 2}\;. \end{aligned} \tag{39}$$

Notice the factor of two difference between this expression and Eq. (31). Although this factor of two might be puzzling initially, it is a straightforward consequence of the fact that H is quadratic in I near $I=0$, i.e., for $I \ll 1$,

$$H=\log D+\frac{\Delta H}{\Delta I}I=\log D-\frac{I^2\ln 2}{2(D-1)}\;, \tag{40}$$

which gives

$$\frac{dH}{dI}=-\frac{I\ln 2}{D-1}\;. \tag{41}$$

In applying these ideas to hypersensitivity of quantum maps, the idea is that D is given roughly by 2^{H_S}, so if H_S increases linearly with time, then D increases exponentially, and you have exponential hypersensitivity to perturbation. The point is that you can have a linear increase of system entropy without having exponential hypersensitivity, because hypersensitivity measures more than the linear entropy increase, having to do with the random distribution of the perturbed vectors in Hilbert space.