

Amplification, Attenuation, and Integration

H. Sebastian Seung

Introduction

Differential equations such as

$$\tau \dot{x}_i + x_i = f\left(\sum_j W_{ij} x_j + b_i\right) \quad (1)$$

have long been used to model networks of interacting neurons (Ermentrout, 1998; PHASE PLANE ANALYSIS OF NEURAL NETS). The activity of neuron i is represented by a single dynamical variable x_i , and its input-output characteristics by a single transfer function f . There are more biophysically realistic descriptions of neural networks that include many dynamical variables per neuron, in order to explicitly model dendritic integration, action potential generation, and synaptic transmission. Nevertheless, simplified models like that in Equation 1 have been useful for understanding how the computational properties of neural networks are related to their synaptic organization.

The parameter W_{ij} in Equation 1 represents the strength of the synapse from neuron j to i . These synapses are termed *recurrent*, as they connect to other neurons in the same network. Feedforward synaptic input from outside the network is implicit in the bias b_i . The feedforward synapses could be made explicit by writing $b_i = b_i^0 + \sum_a V_{ia} z_a$, where z_a are input neuron activities, V_{ia} the strengths of the feedforward synapses, and b_i^0 any intrinsic tendency of neuron i to be active. But the feedforward connections will be left implicit in the following, so as to focus on the computational role of the recurrent connections.

Accordingly, the biases b_i in Equation 1 will be regarded as the inputs to the network, while the activities x_i are the outputs. If there were no recurrent synapses ($W_{ij} = 0$ for all i and j), then each neuron i would respond by low-pass filtering the signal $f(b_i)$ with time constant τ . When there are recurrent synapses, a general characterization of the response properties of a network is difficult, but the situation is greatly simplified when nonlinearity is neglected. Putting the transfer function $f(u) = u$ in Equation 1 yields the linear network

$$\tau \dot{x}_i + x_i = \sum_j W_{ij} x_j + b_i \quad (2)$$

which can be completely analyzed using the tools of linear systems theory. The modest goal of this article is to describe some properties of linear networks and give examples of their application to neural modeling.

In particular, the focus is on the role of recurrent synaptic connectivity. Provided that they do not lead to instability, the recurrent connections alter both the gain and speed of response to feedforward input. Either they amplify and slow down responses to feedforward input, or they attenuate and speed up responses. Both effects can occur simultaneously in the same network, as can be seen by mathematically transforming the network of interacting neurons into a set of noninteracting eigenmodes. The effect of the recurrent synapses generally varies from mode to mode.

Besides amplification and attenuation, a linear network can also carry out the operation of temporal integration, in the sense of Newtonian calculus. This happens when the strength of feedback is precisely tuned for an eigenmode, so that its gain and time constant diverge to infinity.

Admittedly, the neglect of nonlinearity is a step away from biological realism. Nevertheless, linear models are important because they give insight into the local behavior of nonlinear networks, which can often be linearly approximated in the vicinity of fixed points. And the linear computations of amplification, attenuation, and integration have been ascribed to a number of brain areas.

65 **Autapse**

66 The simplest example of a recurrent synapse is a single neuron
 67 with a synapse onto itself, or *autapse*, in the terminology of neu-
 68 rophysiology. For this case, the dynamics (Equation 2) takes the
 69 form

$$70 \quad \tau \dot{x} + x = Wx + b \quad (3)$$

72 The autapse has strength W and is said to be excitatory if $W > 0$
 73 and inhibitory if $W < 0$. The example is not meant to be a realistic
 74 model of a biological autapse; it is only a simple illustration of
 75 some of the effects of recurrent synaptic connections. The param-
 76 eter W will also be called the strength of *feedback*, in the termi-
 77 nology of engineering. Without feedback ($W = 0$), the neuron acts
 78 as a low-pass filter of input b with time constant τ . When the effect
 79 of feedback is considered, the first distinction that has to be made
 80 is between the unstable $W > 1$ and the stable $W < 1$ cases. (Dis-
 81 cussion of the borderline $W = 1$ case is postponed until later.)

82 If $W > 1$, the autapse is unstable, as can be seen by solving
 83 Equation 3 for input b that is constant in time. The solution diverges
 84 exponentially to infinity, because the feedback is so strong that it
 85 leads to runaway instability. Note that in a more realistic model,
 86 the growth of this runaway instability would eventually be limited
 87 by nonlinearity, but in the idealized linear model (Equation 3),
 88 divergence to infinity is possible.

89 If $W < 1$, the autapse is stable, and the dynamics (Equation 3)
 90 can be rewritten in the form

$$91 \quad \frac{\tau}{1 - W} \dot{x} + x = \frac{b}{1 - W} \quad (4)$$

93 From this formula can be read two numbers that characterize the
 94 autapse: the steady-state gain, and the time constant of response.
 95 The gain is operationally defined by holding the input constant and
 96 allowing the output to relax to the steady-state value $x_\infty = b/(1 -$
 97 $W)$. Then the steady-state gain, defined as the ratio of output x_∞ to
 98 input b , is $1/(1 - W)$. By this definition, the gain is exactly unity
 99 in the case of no feedback ($W = 0$). Positive ($W > 0$) and negative
 100 ($W < 0$) feedback have different effects. Positive feedback ampli-
 101 fies, boosting the gain to a value greater than 1. Negative feedback
 102 attenuates, making the gain less than 1.

103 Positive and negative feedback also have opposite effects on the
 104 speed of response. The time constant of the exponential relaxation
 105 to the steady state is $\tau/(1 - W)$. In the case of no feedback, this is
 106 equal to the fundamental time constant τ . But positive feedback
 107 lengthens the time constant, while negative feedback shortens it.
 108 This means that there is a trade-off between amplification and
 109 speed, sometimes known as the gain-bandwidth trade-off. Intui-
 110 tively speaking, the trade-off arises because feedback amplification
 111 requires that the signal circulate in the feedback loop, so that more
 112 amplification requires more time.

113 In summary, a feedback loop containing a perfectly linear ele-
 114 ment behaves in a simple way. Positive feedback ($W > 0$) amplifies
 115 and slows down response, assuming that it doesn't lead to insta-
 116 bility. Negative feedback ($W < 0$) attenuates and speeds up re-
 117 sponse.

118 The idea of amplification by positive feedback has been promi-
 119 nent in a number of models of primary visual cortex (Douglas et
 120 al., 1995). Neurons in layer 4 receive both feedforward drive from
 121 the thalamus and recurrent input from other cortical neurons. It has
 122 been proposed that the recurrent interactions amplify the responses
 123 to feedforward input. To test this idea, Ferster and colleagues re-
 124 corded from layer 4 neurons. They inactivated corticocortical in-
 125 puts both by cooling (Ferster, Chung, and Wheat, 1996) and elec-
 126 trical stimulation (Chung and Ferster, 1998). In both cases, they
 127 measured a two- or threefold reduction in the amplitude of cortical
 128 responses to visual stimulation, which was interpreted as a loss of
 129 amplification by positive feedback.

130 The above discussion omitted the special case of $W = 1$, which
 131 is the borderline between stability and instability. For $W \neq 1$, there
 132 was exactly one steady state, which was either stable or unstable,
 133 depending on whether W was less than or greater than 1. In contrast,

134 if $W = 1$, there is not a unique steady state. The number of steady
 135 states depends on b . There are infinitely many if $b = 0$, and none
 136 at all if $b \neq 0$. To understand the case of non-zero b , it is helpful
 137 to return to Equation 3, which reduces to $\tau\dot{x} = b$. In other words,
 138 the response x is the time integral of b . Therefore, a linear autapse
 139 can act as an integrator, if the strength of feedback is precisely
 140 tuned (Seung et al., 2000). Variants of this idea have been used to
 141 model neural integrators, brain areas that integrate their inputs in
 142 the sense of Newtonian calculus (Robinson, 1989).

143 **Mutually Inhibitory Pair**

144 While the autapse illustrates the gain-bandwidth trade-off in feed-
 145 back amplification, it involves only a single neuron, and cannot
 146 capture genuine population behaviors. A more interesting example
 147 consists of two linear neurons with mutual inhibition:

$$148 \quad \tau\dot{x}_1 + x_1 = -\beta x_2 + b_1 \quad (5)$$

$$150 \quad \tau\dot{x}_2 + x_2 = -\beta x_1 + b_2 \quad (6)$$

152 The parameter β is assumed to be positive, so that the interaction
 153 is inhibitory. This dynamics is more complex than Equation 3 be-
 154 cause it involves two differential equations that are coupled to each
 155 other. Luckily, it turns out that the equations can be decoupled by
 156 adding and subtracting them.

157 Adding the two equations yields an equation for the common
 158 mode $x_c = x_1 + x_2$,

$$159 \quad \tau \frac{d}{dt} (x_1 + x_2) + (x_1 + x_2) = -\beta(x_1 + x_2) + (b_1 + b_2) \quad (7)$$

161 Comparison with Equation 3 reveals that the common mode be-
 162 haves like an autapse with negative feedback. Therefore the com-
 163 mon mode attenuates its input $b_1 + b_2$ with steady-state gain $1/(1$
 164 $+ \beta)$ and time constant $\tau/(1 + \beta)$.

165 Similarly, subtracting the two equations yields an equation for
 166 the differential mode $x_d = x_1 - x_2$,

$$167 \quad \tau \frac{d}{dt} (x_1 - x_2) + (x_1 - x_2) = \beta(x_1 - x_2) + (b_1 - b_2) \quad (8)$$

169 The differential mode behaves like an autapse with positive feed-
 170 back. If $\beta > 1$, the differential mode is unstable. If $\beta < 1$, then the
 171 differential mode amplifies its input $b_1 - b_2$ with steady-state gain
 172 $1/(1 - \beta)$ and time constant $\tau/(1 - \beta)$.

173 To recapitulate, transforming from (x_1, x_2) to (x_c, x_d) formally
 174 decoupled the mutually inhibitory pair of neurons into two “virtual”
 175 autapses. Note that the transformation is reversible, as x_1 and x_2
 176 can be reconstructed from the common and differential modes, e.g.,
 177 $x_1 = (x_c + x_d)/2$.

178 A striking aspect of this example is that mutual inhibition has
 179 completely opposite effects on the common and differential modes.
 180 For the common mode, inhibition mediates negative feedback,
 181 which leads to attenuation. But inhibition mediates positive feed-
 182 back for the differential mode, which leads to amplification.

183 The general lesson to be drawn is that no direct correspondence
 184 exists between the sign of synaptic connections and the sign of
 185 feedback. This is because a synapse is local, belonging to just two
 186 neurons. In contrast, feedback strength is global, belonging to a
 187 distributed mode of the network. As will be described below, the
 188 feedback strength is given in general by the eigenvalues of the
 189 synaptic weight matrix W . The autapse is a special exception for
 190 which the sign of the synaptic connection directly corresponds to
 191 the sign of feedback, but this does not hold true in general.

192 The idea that inhibition between neurons can amplify differences
 193 has been used to explain the fact that visual systems are more
 194 sensitive to relative luminance, or contrast, than to absolute lumi-
 195 nance. For example, the *Limulus* retina consists of visual receptors
 196 that are topographically organized in a two-dimensional network
 197 and interact via lateral inhibition. Measurements of retinal output
 198 reveal enhancement of luminance differences, a fact that has been
 199 successfully explained using network models that are generaliza-
 200 tions of the mutually inhibitory pair considered here (Hartline and
 201 Ratliff, 1972).

202 The special case $\beta = 1$ is also of interest. It is the borderline of

203 stability for the differential mode. If $b_1 - b_2$ is zero, then the
 204 differential mode $x_1 - x_2$ is constant in time, according to Equation
 205 8, while the common mode $x_1 + x_2$ converges exponentially to the
 206 value $(b_1 + b_2)/2$. This is a simple example of a *line attractor*, a
 207 line of fixed points to which all trajectories are attracted (Seung,
 208 1996). More complex nonlinear network models with approximate
 209 line attractors have been used to model the phenomenon of persis-
 210 tent neural activity (Seung, 1996; Zhang, 1996).

211 Note that having a continuous set of fixed points is an unusual
 212 situation, requiring the precise tuning of the inhibitory strength β
 213 and the differential input $b_1 - b_2$. When $b_1 - b_2$ is non-zero, then
 214 it is integrated by the differential mode. In this case, inhibitory
 215 interactions yield an integrator, in contrast to the autapse, which
 216 requires excitatory feedback to integrate. Robinson et al. proposed
 217 that lateral inhibition is the mechanism of the oculomotor neural
 218 integrator, which converts vestibular and other velocity-coded in-
 219 puts into eye position outputs (Cannon, Robinson, and Shamma,
 220 1983).

221 General Network

222 For a general network of N neurons, the effects of feedback can be
 223 understood via eigensystem analysis. It is convenient to rewrite the
 224 dynamics in Equation 2 in matrix-vector form as

$$225 \quad \tau \frac{d}{dt} x + x = Wx + b \quad (9)$$

227 where x and b are vectors and W is the synaptic weight matrix.

228 Suppose that the weight matrix can be factorized as $W = S\Lambda S^{-1}$,
 229 where Λ is a real diagonal matrix and S is a real invertible matrix.
 230 A sufficient condition for a real diagonalization is that the weight
 231 matrix W be symmetric, but this is not a necessary condition. The
 232 diagonal entries of Λ are the eigenvalues of W . The columns of S
 233 are the right eigenvectors of W , and the rows of S^{-1} are the left
 234 eigenvectors.

235 Recall that transforming to the common and differential modes
 236 simplified the dynamics of the mutually inhibitory pair. The ana-
 237 logue here is to change from x and b to

$$238 \quad \tilde{x} = S^{-1}x \quad \tilde{b} = S^{-1}b$$

240 These vectors can be used to express x and b as linear combinations
 241 of the right eigenvectors, $x = S\tilde{x}$ and $b = S\tilde{b}$.

242 The transformation of Equation 9 is effected by multiplying with
 243 S^{-1} ,

$$244 \quad \tau \frac{d}{dt} \tilde{x} + \tilde{x} = S^{-1}Wx + \tilde{b} \quad (10)$$

$$245 \quad = S^{-1}WS\tilde{x} + \tilde{b} \quad (11)$$

$$246 \quad = \Lambda\tilde{x} + \tilde{b} \quad (12)$$

250 Writing out the last expression component by component yields

$$251 \quad \tau \frac{d}{dt} \tilde{x}_a + \tilde{x}_a = \lambda_a \tilde{x}_a + \tilde{b}_a$$

253 where λ_a is the a th diagonal element of Λ , or equivalently the a th
 254 eigenvalue of W . This is a great simplification: the network (Equa-
 255 tion 9) of N interacting neurons has been transformed into N non-
 256 interacting “virtual” autapses. Each autapse has feedback with
 257 strength given by the eigenvalues λ_a . Assuming that the eigenval-
 258 ues are less than or equal to 1, each autapse can perform the opera-
 259 tions of amplification, attenuation, or integration.

260 Discussion

261 In this article, some effects of recurrent synaptic connectivity on
 262 linear networks were characterized. The autapse example demon-
 263 strated that there is a gain-bandwidth trade-off in amplification and
 264 attenuation by feedback, and the possibility of integration when
 265 feedback is precisely tuned. The mutually inhibitory pair illustrated
 266 the decoupling of an interacting network into “virtual” autapses,
 267 and also illustrated that the sign of feedback is not directly related
 268 to the sign of synaptic connections. Such a decoupling is generally

269 possible for any synaptic weight matrix W that is diagonalizable
270 with all real eigenvalues.

271 More generally, the eigenvalues (and eigenvectors) are complex
272 numbers. When an eigenvalue of W has a non-zero imaginary part,
273 the corresponding eigenmode exhibits oscillatory behavior. Ac-
274 cordingly, linear analyses have been used to explain the existence
275 of oscillations in some neural network models (Li and Hopfield,
276 1989).

277 It is natural to ask whether the concepts introduced above have
278 any relevance for *nonlinear* neural networks. A simple way of mod-
279 eling nonlinearity is to introduce a threshold for activation by
280 choosing $f(x) = \max\{x, 0\}$ for the transfer function in Equation 1.
281 Because the resulting dynamics are piecewise linear, eigenvalues
282 and eigenvectors are still essential for mathematical analysis (Had-
283 eler and Kuhn, 1987; Hahnloser et al., 2000), but the threshold
284 nonlinearity leads to a richer variety of dynamical behaviors. A full
285 discussion of threshold linear networks is beyond the scope of this
286 article, but let us briefly reconsider the example of a mutually in-
287 hibitory pair of neurons presented with inputs that are constant in
288 time. For linear neurons, the mutual inhibition caused differences
289 in input to be amplified in the steady-state response. If the neurons
290 are instead threshold linear, *winner-take-all* behavior can result for
291 some choices of model parameters. Then only a single neuron is
292 active at steady state, no matter how small the difference in inputs
293 may be (Amari and Arbib, 1977). As in the purely linear case, the
294 difference in steady-state outputs is greater than the difference in
295 inputs. However, this behavior cannot be explained in terms of a
296 simple linear amplification. For a more detailed explanation, see
297 WINNER-TAKE-ALL NETWORKS.

298 **Roadmap:** Dynamic Systems

299 **Background:** Dynamics and Adaptation in Neural Networks

300 **Related Reading:** Pattern Formation, Neural; Winner-Take-All Networks

301 References

- 302 Amari, S., and Arbib, M. A., 1977, Competition and cooperation in neural
303 nets, in *Systems Neuroscience* (J. Metzler, Ed.), New York: Academic
304 Press, pp. 119–165. ◆
- 305 Cannon, S. C., Robinson D. A., and Shamma, S. 1983, A proposed neural
306 network for the integrator of the oculomotor system, *Biol. Cybern.*,
307 49:127–136.
- 308 Chung, S., and Ferster, D., 1998, Strength and orientation tuning of the
309 thalamic input to simple cells revealed by electrically evoked cortical
310 suppression, *Neuron*, 20:1177–1189.
- 311 Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., and Suarez,
312 H. H., 1995, Recurrent excitation in neocortical circuits, *Science*,
313 269:981–985.
- 314 Ermentrout, B., 1998, Neural networks as spatio-temporal pattern-forming
315 systems, *Rep. Prog. Phys.*, 61:353–430. ◆
- 316 Ferster, D., Chung, S., and Wheat, H., 1996, Orientation selectivity of tha-
317 lamic input to simple cells of cat visual cortex, *Nature*, 380(6571):249–
318 252.
- 319 Hadeler, K. P., and Kuhn, D., 1987, Stationary states of the Hartline-Ratliff
320 model, *Biol. Cybern.*, 56:411–417.
- 321 Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., and
322 Seung, H. S., 2000, Digital selection and analogue amplification coexist
323 in a cortex-inspired silicon circuit, *Nature*, 405(6789):947–951.
- 324 Hartline, H. K., and Ratliff, F., 1972, Inhibitory interaction in the retina of
325 *Limulus*, in *Handbook of Sensory Physiology: Physiology of Photore-
326 ceptor Organs* (M. G. F. Fuortes, Ed.), Berlin: Springer-Verlag, pp. 382–
327 447.
- 328 Li, Z., and Hopfield, J. J., 1989, Modeling the olfactory bulb and its neural
329 oscillatory processings, *Biol. Cybern.*, 61:379–392.
- 330 Robinson, D. A., 1989, Integrating with neurons, *Annu. Rev. Neurosci.*,
331 12:33–45.
- 332 Seung, H. S., 1996, How the brain keeps the eyes still, *Proc. Natl. Acad.
333 Sci. USA*, 93:13339–13344.
- 334 Seung, H. S., Lee, D. D., Reis, B. Y., and Tank, D. W., 2000, The autapse:
335 A simple illustration of short-term analog memory storage by tuned syn-
336 aptic feedback, *J. Comput. Neurosci.*, 9:171–185. ◆
- 337 Zhang, K., 1996, Representation of spatial orientation by the intrinsic dy-
338 namics of the head-direction cell ensemble: A theory, *J. Neurosci.*,
339 16:2112–2126.
- 340