CNS 187 Neural Computation - Problem Sheet 3

Ling Li, ling@cs.caltech.edu

October 24, 2000

## 3.1 Lyapunov Energy Functions

1. The damping force is $-\gamma\dot{x}$, and the nonlinear spring force is $-\kappa x(1+\delta x^2)$. Thus the dynamic equation is
$$m\ddot{x} = -\kappa x(1+\delta x^2) - \gamma\dot{x}. \tag{1}$$

2. The total energy of the system is sum of the kinetic energy and the elastic energy:
$$E = \frac{1}{2}m\dot{x}^2 + \int_0^x \kappa s(1+\delta s^2)ds = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}\kappa x^2 + \frac{1}{4}\kappa\delta x^4. \tag{2}$$

3. To find the steady-state of the system, we substitute $\dot{x}$ and $\ddot{x}$ in (1) with 0 and get $x = 0$. Thus $x = \dot{x} = \ddot{x} = 0$ is the only steady-state. From (2), we get
$$\begin{aligned}\frac{\partial E}{\partial t} &= m\dot{x}\ddot{x} + \kappa x\dot{x} + \kappa\delta x^3\dot{x} \\ &= \dot{x}\left[-\kappa x(1+\delta x^2) - \gamma\dot{x}\right] + \kappa x\dot{x}(1+\delta x^2) \\ &= -\gamma\dot{x}^2 \le 0\end{aligned}$$
with equality iff $\dot{x} = 0$. With $\kappa > 0$, $\delta > 0$ and obviously $m > 0$, we have energy $E$ in (2) is non-negative, thus bounded below. So the energy of the system will always decrease until $\frac{\partial E}{\partial t} = 0$ for $t$ bigger than some $t_0$, which leads to $\ddot{x} = 0$ and $x = 0$. So the steady state $x = \dot{x} = \ddot{x} = 0$ is stable.

## 3.2 Lyapunov Analysis of Networks

Assume $g(\cdot)$ is bounded above and below. That is, there exists a constant $M$ such that for all $x$
$$|g(x)| \le M.$$
Also assume $g'(\cdot) > 0$, $\tau_i > 0$, and $I_i$ is constant.

1. We argue that $\alpha_i\beta_i > 0$ is sufficient for $\mathcal{L}$ to be a Lyapunov function for the network.

   - **$\mathcal{L}$ is bounded below**. From
$$S_i = \alpha_i g(\beta_i V_i) \tag{3}$$

we get

$$\frac{dS_i}{dV_i} = \alpha_i \beta_i g'(\beta_i V_i) > 0.$$

Since $G_i(S_i) = G_i(g_i(V_i))$ is a function of $V_i$,

$$\frac{dG_i(S_i)}{dV_i} = G_i'(S_i)\frac{dS_i}{dV_i} = \alpha_i \beta_i g'(\beta_i V_i) V_i,$$

which has the same sign as $V_i$. Thus $G_i(S_i)$ has its minimum value at $V_i = 0$, i.e.,

$$G_i(S_i) \geq G_i(g_i(0)).$$

From the boundedness of $g(\cdot)$, we also know

$$-\frac{1}{2}\sum_{i,j} w_{ij} S_i S_j - \sum_i S_i I_i \geq -\frac{1}{2}\left(\sum_{i,j}|w_{ij}|\right) M^2 - \left(\sum_i |I_i|\right) M.$$

So

$$\mathcal{L} \geq -\frac{1}{2}\left(\sum_{i,j}|w_{ij}|\right) M^2 + \sum_i G_i(g_i(0)) - \left(\sum_i |I_i|\right) M,$$

which means $\mathcal{L}$ is bounded below.

- $\frac{d\mathcal{L}}{dt} \leq 0$ **with equality iff** $\dot{V}_i = 0$ **for all** $i$. For a symmetric network,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial V_i} &= -\sum_j w_{ij} S_j \frac{dS_i}{dV_i} + G_i'(S_i)\frac{dS_i}{dV_i} - I_i \frac{dS_i}{dV_i} \\
&= -\left(\sum_j w_{ij} S_j - V_i + I_i\right)\frac{dS_i}{dV_i} \\
&= -\tau_i \alpha_i \beta_i g'(\beta_i V_i) \dot{V}_i.
\end{aligned}$$

Thus

$$\frac{d\mathcal{L}}{dt} = \sum_i \frac{\partial \mathcal{L}}{\partial V_i}\dot{V}_i = -\sum_i \tau_i \alpha_i \beta_i g'(\beta_i V_i)\dot{V}_i^2 \leq 0$$

with equality iff $\dot{V}_i = 0$ for all $i$.

2. When $r = 1$, we have $C_{ij}^r = C_{ij} = C_{ji}^r$ for $i \neq j$ and $C_{ii}^r = 0$. That is a symmetric recurrent network. When $r$ increases, the network becomes an asymmetric one, and the connections from neurons with bigger numbers to neurons with smaller numbers are weaker than those from neurons with smaller numbers to neurons with bigger numbers. In other words, the forward connections are dominant while the backward connections are decaying. Finally, $r$ goes to $\infty$, there's no backward connections and the network is now a feedforward network.

3. The equation for the asymmetric recurrent network is

$$\tau \dot{u}_i = -u_i + \sum_j C_{ij}^r g(u_j) + i_i. \tag{4}$$

Assuming $u_i = \beta_i V_i$ ($\beta_i \neq 0$), we get

$$\tau \beta_i \dot{V}_i = -\beta_i V_i + \sum_j C^r_{ij} g(\beta_j V_j) + i_i.$$

That is

$$\tau \dot{V}_i = -V_i + \sum_j \frac{C^r_{ij}}{\beta_i \alpha_j} \alpha_j g(\beta_j V_j) + \frac{i_i}{\beta_i}.$$

Compared to (3) and the equation for the symmetric network

$$\tau_i \dot{V}_i = -V_i + \sum_{j=1} w_{ij} S_j + I_i, \tag{5}$$

we get the equivalence of (4) and (5) requires

$$\tau_i = \tau, I_i = \frac{i_i}{\beta_i}, \tag{6}$$

and

$$w_{ij} = \frac{C^r_{ij}}{\beta_i \alpha_j} = \begin{cases} \dfrac{C_{ij}}{\beta_i \alpha_j}, & \text{if } i > j; \\ 0, & \text{if } i = j; \\ \frac{C_{ji} r^{i-j}}{\beta_i \alpha_j}, & \text{if } i < j. \end{cases}$$

In (5), $w_{ij} = w_{ji}$. Thus for $i > j$,

$$\frac{C_{ij}}{\beta_i \alpha_j} = \frac{C_{ij} r^{j-i}}{\beta_j \alpha_i}$$

We should ensure that

$$\beta_i \alpha_j = r^{i-j} \beta_j \alpha_i,$$

which can be rewritten as (since $\alpha_i \beta_i > 0$ and $r > 0$)

$$r^{-i} \alpha_i^{-1} \beta_i = r^{-j} \alpha_j^{-1} \beta_j > 0.$$

That is, for all $i$, $r^{-i} \alpha_i^{-1} \beta_i$ should be a positive constant. Let it be $\xi > 0$. Thus

$$\alpha_i = r^{-i} \beta_i / \xi, \tag{7}$$

$$w_{ij} = \begin{cases} \xi \dfrac{C_{ij} r^j}{\beta_i \beta_j}, & \text{if } i > j; \\ 0, & \text{if } i = j; \\ \xi \frac{C_{ji} r^i}{\beta_i \beta_j}, & \text{if } i < j. \end{cases} \tag{8}$$

Thus given $\tau$, $C_{ij}$, $\beta_i$, we can first choose $\xi > 0$, then use (6), (7) and (8) to make the symmetric network (5) equivalent to (4).

4. The equations for the network with 3 nodes are

$$\begin{aligned} \tau \dot{V}_1 &= -V_1 - \tanh(\beta V_3) \\ \tau \dot{V}_2 &= -V_2 - \tanh(\beta V_1) \\ \tau \dot{V}_3 &= -V_3 - \tanh(\beta V_2) \end{aligned}$$

At the steady state, $\dot{V}_1 = \dot{V}_2 = \dot{V}_3 = 0$. Thus

$$V_1 = -\tanh(\beta V_3) = \tanh(\beta \tanh(\beta V_2)) = -\tanh(\beta \tanh(\beta \tanh(\beta V_1))).$$

Whether $\beta = 0$ or not, we have $V_1 = 0$, thus $V_2 = V_3 = 0$. When $V_1 = V_2 = V_3 = 0$, we also have $\dot{V}_1 = \dot{V}_2 = \dot{V}_3 = 0$. So, $V_1 = V_2 = V_3 = 0$ is the unique steady state of the network.
For small $x$, we have $\tanh(\beta x) = \beta x + o(x)$. So when there's a small disturbance in state $V_1 = V_2 = V_3 = 0$, we have

$$\begin{aligned} \tau \dot{V}_1 &= -V_1 - \beta V_3 - o(V_3) \\ \tau \dot{V}_2 &= -V_2 - \beta V_1 - o(V_1) \\ \tau \dot{V}_3 &= -V_3 - \beta V_2 - o(V_2) \end{aligned}$$

Let $\mathbf{V} = \begin{pmatrix} V_1 & V_2 & V_3 \end{pmatrix}^T$, we get

$$\dot{\mathbf{V}} = -\frac{1}{\tau} \begin{pmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{pmatrix} \mathbf{V} - o(\mathbf{V})$$

The eigenvalues of $-\frac{1}{\tau} \begin{pmatrix} 1 & 0 & \beta \\ \beta & 1 & 0 \\ 0 & \beta & 1 \end{pmatrix}$ is

$$\lambda_k = -\frac{1}{\tau}\left(1 + \beta e^{\frac{2k\pi}{3} i}\right) = -\frac{1}{\tau}\left(1 + \beta \cos\frac{2k\pi}{3} + \beta i \sin\frac{2k\pi}{3}\right),$$

where $k = 0, 1, 2$. Since the close form of $\mathbf{V}$ includes item $e^{\lambda_k t}$, the convergence of $\mathbf{V}$ requires

$$1 + \beta \cos\frac{2k\pi}{3} > 0$$

for all $k = 0, 1, 2$, which is just $\beta < 2$. Thus $\beta_0 = 2$, and for $\beta \geq \beta_0$, the network will oscillate.

### 3.3 Computing by Optimization: Graph Bipartitioning

Assume the total number of edges and vertices in the graph is $M$ and $N$ ($N$ even), respectively, and the maximum degree of vertices is $d_{\max}$. We use *balanced partitioning* to denote a partitioning with equal number of vertices in each group; and *unbalanced partitioning* when the numbers of vertices in those two groups are not equal.

1. Using notations in the problem, the number of edges between those two groups is

$$\mathcal{L}_M = \frac{1}{2}\sum_{i,j} C_{ij}\left(\frac{1 - S_i S_j}{2}\right) = \frac{M}{2} - \frac{1}{2}\sum_{i,j}\frac{C_{ij}}{2} S_i S_j.$$

Our goal is to minimize $\mathcal{L}_M$ while subject to the balance constraint

$$\sum_i S_i = 0. \tag{9}$$

We can define an energy item for (9)

$$\mathcal{L}_N = \eta \left( \sum_i S_i \right)^2 ,$$

thus we can try to construct energy function like

$$\mathcal{L} = \mathcal{L}_M + \mathcal{L}_N = \frac{M}{2} - \frac{1}{2} \sum_{i,j} \frac{C_{ij}}{2} S_i S_j + \eta \left( \sum_i S_i \right)^2 ,$$

which can be rewritten (with constant $M$ omitted) as

$$\mathcal{L} = -\frac{1}{2} \sum_{i,j} \left( \frac{C_{ij}}{2} - 2\eta \right) S_i S_j. \tag{10}$$

(10) means

$$w_{ij} = \frac{C_{ij}}{2} - 2\eta, I_i = 0, \tag{11}$$

in the Lyapunov function.[1]
To ensure the global minimum of the network is a solution to the problem, we need

$$\eta > \frac{d_{\max}}{4}. \tag{12}$$

**Lemma:** Global minimum is a balanced partitioning.

**Proof:** Suppose (9) doesn't hold at the global minimum. Without loss of generality, we can suppose that there are $2k$ ($k \geq 1$) more vertices in group one than in group two. We can move $k$ arbitrary vertices in group one to group two, trying to balance this two groups, and calculate the change of energy:

$$\begin{aligned} \Delta\mathcal{L} &= \Delta\mathcal{L}_M + \Delta\mathcal{L}_N = \Delta\mathcal{L}_M - 4k^2\eta \\ &< k d_{\max} - k^2 d_{\max} \leq 0. \end{aligned}$$

This contradicts the definition of 'global minimum.'

**Theorem:** Global minimum is a solution to the graph bipartitioning problem.

**Proof:** If a solution with energy $\tilde{\mathcal{L}}$ has less edges between the two groups than the global minimum with energy $\mathcal{L}$, thus means $\tilde{\mathcal{L}}_M < \mathcal{L}_M$ and $\tilde{\mathcal{L}}_N = 0$. From the lemma, $\mathcal{L}_N = 0$. So $\tilde{\mathcal{L}} < \mathcal{L}$, which contradicts the definition of 'global minimum.'

Thus, with (11) and (12), the global minimum of Lyapunov function (10) is a solution to the graph bipartitioning problem.

2. From the proof of the lemma, we know that changing the sign of any single $S_i$ in a balanced partitioning results an unbalanced partitioning and thus increases the energy $\mathcal{L}$. So each balanced partitioning is a local minimum of $\mathcal{L}$. Consider the bipartitioning $\{1,2,3,4,6\}$-$\{5,7,8,9,10\}$. This is balanced so it is a local minimum, but not a global minimum, since $\{1,2,3,4,5\}$-$\{6,7,8,9,10\}$ is better than it.

[1] When $j = i$, $\left( \frac{C_{ij}}{2} - 2\eta \right) S_i S_j = \frac{C_{ii}}{2} - 2\eta$ is a constant. So we can also define $w_{ii} = 0$.

3. With our settings, any balanced partitioning is a local minimum. And, for any unbalanced partitioning, we can change one single $S_i$ in the group with more vertices to reduce the energy. Thus the number of local minima is exactly[2]

$$\binom{N}{\frac{N}{2}} \geq \frac{2^N}{N}.$$

4. We know this is an NP-Complete problem, so intuitively the network cannot do better than other methods. I've thought of two possible situations:

   (a) When gain increases, the distribution of local minima changes and the global minimum at a lower gain becomes a non-global local minimum, and the network can not jump out of that non-global local minimum. Thus the global minimum is not guaranteed by gradually raising the gain.

   (b) The network may find the global minimum. However, that requires the change of gain should be very very small, which takes fairly long time to get to the global minimum. That 'fairly long time' is an exponential of the problem size.

   However, I cannot prove either of them.

[2]The reason for the inequality is

$$\binom{N}{\frac{N}{2}} = 2^2 \frac{N-1}{N} \binom{N-2}{\frac{N}{2}-1} = \ldots = 2^N \frac{N-1}{N} \frac{N-3}{N-2} \cdots \frac{1}{2} \geq \frac{2^N}{N}$$