Sparse approximation via locally competitive algorithms

AMSC PhD Preliminary Exam

Matthew Guay

May 14, 2014

Norbert Wiener Center
for Harmonic Analysis and Applications

Outline I

Table of Contents

Mathematical motivation

- The challenge: How can we (efficiently) represent and find structure in high-dimensional data sets?
- One potential solution: Look for low-dimensional approximations. For linear signal models, this is the **sparse representation problem**: decomposing a signal $\mathbf{y} \in \mathbb{R}^M$ as $\sum_{n=1}^{N} a_n \varphi_n$, where $\{\varphi_n\}$ forms a finite frame for $\mathbb{R}^M$ and $(a_1, \ldots, a_n)$ has small support.
- Efficiently solving this problem requires computing minimizers of regularized least-squares problems.

Neuroscientific motivation

- The **sparse coding hypothesis** (Olshausen 2004):

  > *Information [carried within neural networks] is represented by a relatively small number of simulataneously active neurons out of a large population.*

- Sparse neural activity is observed in connection with competitive inhibition activity in sensory processing centers across many animal species.
- Excitatory neurons race to fire before broadly-tuned feedback inhibitory signals silence excitatory clusters.

Neuroscientific motivation

- Recent research (Lin 2014) produced strong evidence that in the olfactory processing system of the fruit fly (*Drosophila melanogaster*):
  - Feedback inhibition is responsible for sparse neural activity in Kenyon cells.
  - Sparse Kenyon cell activity is important for proper odor discrimination.
- How does this behavior relate to dynamics and frame theory?

Locally competitive algorithms

- A Hopfield network (Hopfield 1982) is a (continuous or discrete) dynamical system for which a Lyapunov function relates system evolution to the geometry of a corresponding energy surface.
- **Locally competitive algorithms** (LCA) are continuous Hopfield(-like) networks for which the optimization problems of sparse representation are (weak) Lyapunov functions. Further analysis demonstrates robust global convergence properties.
- LCA is simultaneously a linear-nonlinear neural network model with competitive inhibition and an efficient optimization algorithm for sparse representation problems.

Table of Contents

Problem setup

- We wish to approximate a signal $\mathbf{y} \in \mathbb{R}^M$ via synthesis in a unit-norm frame Φ, given measurements $\Phi^T \mathbf{y}$.
- The synthesis operator $\mathcal{S}$ for Φ maps a coefficient vector $\mathbf{a} \in \mathbb{R}^N$ to $\sum_{n=1}^{N} a_n \boldsymbol{\varphi}_n \in \mathbb{R}^M$.
- For finite frames, $\mathcal{S}$ can be represented as a matrix $[\boldsymbol{\varphi}_1 \cdots \boldsymbol{\varphi}_n] \in \mathbb{R}^{M \times N}$. Abuse notation, call this Φ as well.
- Variational methods applied to the space of coeffcents $\mathbb{R}^N$ allow us to solve the synthesis problem, given only analytic information.

Frames and optimization

- For well-behaved functions $\mathbf{C} : \mathbb{R}^N \to \mathbb{R}$, the vector $\mathbf{a}$ can be chosen as the solution of a constrained optimization problem

$$\mathbf{a}^* = \arg\min_{\mathbf{a}\in\mathbb{R}^N} \mathbf{C}(\mathbf{a}) \quad \text{s.t.} \quad \frac{1}{2}||\Phi\mathbf{a} - \mathbf{y}\,||_2^2 < \epsilon$$

- For some choice of $\lambda > 0$, this is equivalent to the unconstrained optimization problem

$$\mathbf{a}^* = \arg\min_{\mathbf{a}\in\mathbb{R}^N} \frac{1}{2}||\Phi\mathbf{a} - \mathbf{y}\,||_2^2 + \lambda\mathbf{C}(\mathbf{a}) \tag{1}$$

- The sparse representation problem: Can $\mathbf{C}(\mathbf{a})$ be chosen so that $\mathbf{a}^*$ has small support?

Sparse representation

- The sparsity of a vector $\mathbf{a} \in \mathbb{R}^N$ is characterized by the ℓ^0 "norm" $||\mathbf{a}||_0 \triangleq |\operatorname{supp}(\mathbf{a})|$.
- Solving
$$\mathbf{a}^* = \arg\min_{\mathbf{a}\in\mathbb{R}^N} \frac{1}{2}||\Phi\mathbf{a} - \mathbf{y}\,||_2^2 + \lambda||\mathbf{a}||_0$$
is an NP-hard problem, computationally intractable.
- Alternate choices of $\mathbf{C}(\mathbf{a})$ in (1) provide tractable alternative problems with the same solution, for certain classes of signal $\mathbf{y}$ and frame Φ.
- Example: Basis pursuit denoising (BPDN)
$$\mathbf{a}^* = \arg\min_{\mathbf{a}\in\mathbb{R}^N} \frac{1}{2}||\Phi\mathbf{a} - \mathbf{y}\,||_2^2 + \lambda||\mathbf{a}||_1 \qquad \text{(BPDN)}$$

A broader class of problems

- We will focus on problems of the form

$$\mathbf{a}^* = \arg\min_{\mathbf{a}\in\mathbb{R}^N} V(\mathbf{a}) \triangleq \frac{1}{2}||\Phi\mathbf{a} - \mathbf{y}\,||_2^2 + \lambda \sum_{n=1}^{N} C(a_n) \qquad (2)$$

- Admissible cost functions C are defined by a differential equation in a later slide. This class of functions includes the BPDN regularization term $||\mathbf{a}||_1$ and several other regularization functions of interest (Charles, 2012).

Admissible cost functions

- All admissible C satisfy:
  - $C(0) = 0$ and $C(x) \geq 0 \ \forall x \in \mathbb{R}$.
  - $C \in C^1((-\infty, 0) \cup (0, \infty))$.
  - C is nonincreasing on $(-\infty, 0)$ and nondecreasing on $(0, \infty)$.

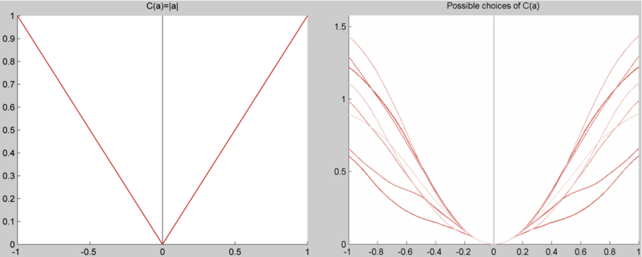


Table of Contents

LCA setup

- Given $\mathbf{y} \in \mathbb{R}^M$ and $\Phi \in \mathbb{R}^M \times N$, the LCA system consists of N "nodes" with internal states $\mathbf{u}(t) : \mathbb{R} \to \mathbb{R}^N$ and output states $\mathbf{a}(t) : \mathbb{R} \to \mathbb{R}^N$.
- Internal and output states related component-wise by a thresholding function $T_\lambda : \mathbb{R} \to \mathbb{R}$ such that $T_\lambda(u_n) = a_n$. Define $\mathbf{T}_\lambda(\mathbf{u}) = (T_\lambda(u_1), \ldots, T_\lambda(u_N))$.
- Given $\mathbf{T}_\lambda$ and a time constant $\tau > 0$, the LCA system is

$$\begin{aligned} \tau\dot{\mathbf{u}} &= -\mathbf{u} - (\Phi^T\Phi - I)\mathbf{a} + \Phi^T\mathbf{y} \\ \mathbf{a} &= \mathbf{T}_\lambda(\mathbf{u}), \end{aligned} \tag{LCA}$$

The function T_λ

- Convergence results in (Balavoine 2012) are established for LCA systems with admissible thresholding functions

$$T_\lambda(u_n) = \begin{cases} 0, & |u_n| \leq \lambda \\ f(u_n), & |u_n| > \lambda \end{cases}$$

- $f : \mathbb{R} \to \mathbb{R}$ is C^1 on $\mathcal{D} = (-\infty, -\lambda] \cup [\lambda, \infty)$ and satisfies

$$\begin{array}{ll} f(-u_n) = -f(u_n) & \forall u_n \in \mathcal{D},\ \ f(\lambda) = 0 \\ f'(u_n) > 0 & \forall u_n \in \mathcal{D} \\ f(u_n) \leq u_n & \forall u_n \in (\lambda, \infty) \end{array}$$

LCA dynamics

- To understand how the LCA works, examine a single node's dynamics:

$$\dot{u}_n = -u_n - \sum_{k \neq n} \langle \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_k \rangle a_k + \langle \boldsymbol{\varphi}_n, \mathbf{y} \rangle$$
$$a_n = T_\lambda(u_n)$$

- Internal state u_n receives input $\langle \boldsymbol{\varphi}_n, \mathbf{y} \rangle$, "loses charge" in the absence of stimulus ($-\mathbf{u}$ term), and receives feedback inhibition (and excitation) from other active, correlated states.
- Output state a_n is rectified by a thresholding function that does not saturate (radially unbounded).

Connections to optimization

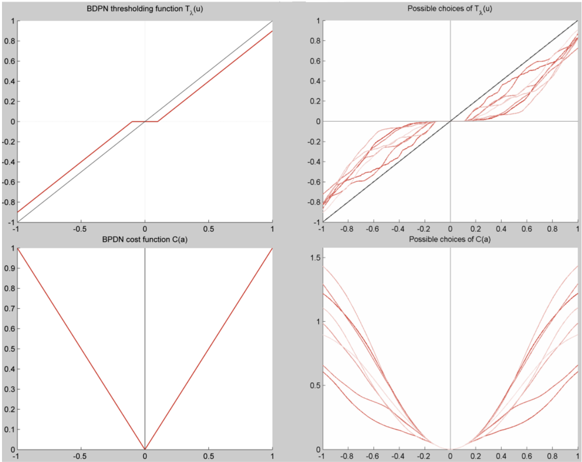
- Define a function $C(a_n) : \mathbb{R} \to \mathbb{R}$ as $C(0) = 0$, and satisfying

$$\lambda \frac{dC}{da_n} = u_n - T_\lambda(u_n) \tag{3}$$

  for $a_n \neq 0$.

- For an admissible T_λ, these functions C are admissible in the sense discussed for cost functions.
- For such cost and thresholding functions, the function $V(\mathbf{a})$ from (2) satisfies $\frac{dV}{da_n} = -\frac{du_n}{dt}$ and is a weak Lyapunov function for (LCA).

Thresholding and cost function illustrations



Connections to optimization

- A function $J : \mathbb{R}^N \to \mathbb{R}$ is a weak Lyapunov function if
  (1) $J(\mathbf{x}) > 0 \ \forall \mathbf{x} \neq \mathbf{0}$;
  (2) J is continuous on $\mathbb{R}^N$;
  (3) $\dot{J}(\mathbf{x}(t)) \leq 0 \ \forall t > 0, \ \mathbf{x} \in \mathbb{R}^N$.
  (4) $\lim_{||\,\mathbf{x}\,|| \to \infty} J(\mathbf{x}) = \infty$.
- $V(\mathbf{a})$ is a weak Lyapunov function for the LCA system: conditions (1),(2),(4) are obvious and (model argument)

$$\dot{V}(\mathbf{a}(t)) = \left\langle \nabla_{\mathbf{a}} V, F' \dot{u} \right\rangle = - \left\langle \dot{u}, F' \dot{u} \right\rangle \leq 0$$

for an $N \times N$ diagonal matrix F' with $F'_{kk} = f'(a_k)$.

Connections to optimization

- LCA trajectories have non-increasing $V(\mathbf{a})$ values in time, but this does not prove global convergence.
- The LCA thresholding function structure violates key assumptions (nonlinear, non-smooth, unbounded) required for global convergence results in past literature.
- The network interconnection matrix $\Phi^T\Phi - I$ may have a non-trivial null-space and/or negative eigenvalues, further complicating analysis.

Active and inactive sets

- Key observation: the LCA system decomposes into active and inactive sets at each time t, based on which output nodes are thresholded.
- For each subset $\Gamma \subseteq \{1, \ldots, N\}$, define $\mathbf{u}_\Gamma$ and $\mathbf{a}_\Gamma$ as the elements of $\mathbf{u}$, $\mathbf{a}$ indexed by Γ. Let Φ_Γ be the matrix composed of columns of Φ indexed by Γ. Same for F'_Γ.
- At each $t \in \mathbb{R}$, $\mathbf{u}(t)$ decomposes into an active set $\mathbf{u}_\Gamma(t)$ and inactive set $\mathbf{u}_{\Gamma^C}(t)$: $|u_k(t)| > \lambda \; \forall\, k \in \Gamma$. $\mathbf{a}(t)$ is decomposed likewise.

LCA as a switched system

- At switching times $\{t_k\}_{k=1}^{\infty}$, the membership of the active set $\Gamma(t)$ changes. For all $k > 0$ and all $s \in [t_{k-1}, t_k)$, $\Gamma(s)$ is some fixed set Γ_k.
- LCA can be recast as a switched system: between switching times, rewrite (LCA) as

$$\dot{\mathbf{a}}_\Gamma(t) = F'_\Gamma(t)\dot{\mathbf{u}}_\Gamma(t) \tag{4}$$

$$\dot{\mathbf{u}}_\Gamma(t) = -\mathbf{u}_\Gamma(t) + \mathbf{a}_\Gamma(t) - \Phi_\Gamma^T \Phi_\Gamma \mathbf{a}_\Gamma + \Phi_\Gamma^T \mathbf{y} \tag{5}$$

$$\dot{\mathbf{u}}_{\Gamma^C}(t) = -\mathbf{u}_{\Gamma^C}(t) - \Phi_{\Gamma^C}^T \Phi_\Gamma \mathbf{a}_\Gamma(t) + \Phi_{\Gamma^C}^T \mathbf{y}\,. \tag{6}$$

Table of Contents

Background - stability and convergence

- Some basic stability results for dynamical systems

$$\dot{\mathbf{x}} = G(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^N . \tag{7}$$

- Fixed point - $\mathbf{x}^* \in \mathbb{R}^N$ such that $G(\mathbf{x}^*) = 0$.
- (Lyapunov) stability - (7) is stable at $\mathbf{x}^*$ if for each $\epsilon > 0$, there exists $R > 0$ such that for all $\mathbf{x}_0$ with $||\,\mathbf{x}_0 - \mathbf{x}^*\,|| < R$, all trajectories $\mathbf{x}(t)$ of (7) with $\mathbf{x}(0) = \mathbf{x}_0$ satisfy $||\,\mathbf{x}(t) - \mathbf{x}^*\,|| < \epsilon$, $\forall t > 0$.
- Asymptotic stability - (7) is asymptotically stable at $\mathbf{x}^*$ if there exists $R > 0$ such that for all $\mathbf{x}_0$ with $||\,\mathbf{x}_0 - \mathbf{x}^*\,|| < R$, trajectories $\mathbf{x}(t)$ of (7) with $\mathbf{x}(0) = \mathbf{x}_0$ satisfy $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$.

Background - stability and convergence

- Global convergence - (7) is globally convergent at $\mathbf{x}^*$ if it is asymptotically stable with $R = \infty$.
- **Global quasi-convergence** - Given $\mathbf{x}(t)$ as in (7), let $\mathbf{z}(t) = T_\lambda(\mathbf{x}(t))$ and let $\mathcal{E} = \{\mathbf{z}^* \in \mathbb{R}^N \mid \dot{\mathbf{z}}^* = 0\}$. (7) is globally quasi-convergent if for all initial $\mathbf{x}_0 \in \mathbb{R}^N$ of (7), $\lim_{t\to\infty} \mathbf{z}(t) \in \mathcal{E}$.
- Most generally, LCA systems will be globally quasi-convergent, with global convergence under additional assumptions.
- Global quasi-convergence for LCA means $\mathbf{a}(t)$ will converge to a fixed point $\mathbf{a}^*$ (possibly non-unique).

Background - subgradients

- For the cost functions C of interest, $V(\mathbf{a})$ is not a differentiable function.
- To study convergence, we need subgradients - $\partial g(\mathbf{x}) = \nabla g(\mathbf{x})$ when the gradient exists, and otherwise $\partial g(\mathbf{x})$ is the convex hull of the set of limit points of $\nabla h(\mathbf{x}_i)$ as $\mathbf{x}_i \to \mathbf{x}$.

$$\partial g(\mathbf{x}) = co\left\{ \lim_{i\to\infty} \nabla g(\mathbf{x}_i) \mid \mathbf{x}_i \to \mathbf{x}, \mathbf{x}_i \notin \Omega_g \right\},$$

where Ω_g is the set of points of nondifferentiability of g.

Background - subgradients

- Critical point - $\mathbf{x}^*$ is a critical point for a nonsmooth function g iff $0 \in \partial g(\mathbf{x}^*)$. When g is convex, all such critical points are minima of g.
- Regularity - $V(\mathbf{a})$ and $C(a_n)$ are regular functions - their one-sided directional derivatives exist at all points in their domain. Therefore we have that

$$\partial V(\mathbf{a}(t)) = -\Phi^T \mathbf{y} + \Phi^T \Phi \mathbf{a}(t) + \lambda \partial C(\mathbf{a}(t)) \qquad (8)$$

- **Lemma 1** (Generalized chain rule): For $V : \mathbb{R}^N \to \mathbb{R}$ regular and $\mathbf{a}(t) : [0, \infty) \to \mathbb{R}^N$ continuously differentiable,

$$\frac{d}{dt} V(\mathbf{a}(t)) = \zeta^T \dot{\mathbf{a}}(t) \qquad \forall \zeta \in \partial V(\mathbf{a}(t)). \qquad (9)$$

LCA convergence overview

- Theorem 1: LCA fixed points are critical points of $V(\mathbf{a})$, trajectories $\mathbf{a}(t)$ are globally quasi-convergent.
- Theorem 2: LCA active sets stop switching after finitely many switches as long as $u_n^* \neq \lambda$ for any n.
- Theorem 3: Given RIP-like conditions on Φ and bounds on $|f'|$, LCA trajectories are globally exponentially convergent to a unique equilibrium.

Theorem 1

- Three main results for the LCA system, assuming admissible cost functions $C(a_n)$ and thresholding functions $f(u_n)$:
  (a) Fixed points of (LCA) are critical points of the objective function $V(\mathbf{a})$.
  (b) LCA $\mathbf{a}(t)$ trajectories are globally quasi-convergent (to critical points of $V(\mathbf{a})$).
  (c) If the critical points of $V(\mathbf{a})$ are isolated, there is a unique critical point and LCA $\mathbf{u}(t)$ are globally convergent.

Part (a) proof sketch

- Want to show: Any fixed point $\mathbf{u}^*$ of (LCA) with active set Γ_* is a critical point of $V(\mathbf{a})$.
- Critical points $\mathbf{a}^*$ of $V(\mathbf{a})$ satisfy

$$0 \in -\Phi^T \mathbf{y} + \Phi^T \Phi \mathbf{a}^* + \lambda \partial \mathbf{C}(\mathbf{a}^*) \tag{10}$$

- $\dot{\mathbf{u}} = 0$ at $\mathbf{u}^*$ can be written separately for active and inactive sets as

$$-\mathbf{u}^*_{\Gamma_*} + f(\mathbf{u}^*_{\Gamma_*}) - \Phi^T_{\Gamma} \Phi_{\Gamma} \mathbf{a}^*_{\Gamma_*} + \Phi^T_{\Gamma_*} \mathbf{y} = 0 \tag{11}$$

$$-\mathbf{u}^*_{\Gamma^C_*} - \Phi^T_{\Gamma^C_*} \Phi_{\Gamma_*} \mathbf{a}^*_{\Gamma_*} + \Phi^T_{\Gamma^C_*} \mathbf{y} = 0 \tag{12}$$

Part (a) proof sketch

- For $\mathbf{u}^*_{\Gamma_*}$, the subgradient in (10) is $\nabla C(\mathbf{a}^*_{\Gamma_*}) = u^*_{\Gamma_*} - f(u^*_{\Gamma_*})$ and (10) is equivalent to (11).
- For $\mathbf{u}^*_{\Gamma^C_*}$, first compute $\lim_{a\downarrow 0} \nabla C(a) = 1$ and $\lim_{a\uparrow 0} \nabla C(a) = -1$, so

$$\partial C(0) = [-1, 1].$$

- Given this, (10) becomes

$$\Phi^T_{\Gamma^C_*} \mathbf{y} - \Phi^T_{\Gamma^C_*} \Phi_{\Gamma_*} \mathbf{a}^*_{\Gamma_*} \in [-\lambda, \lambda],$$

a condition satisfied by (12).

Part (b) proof sketch

- Want to show: LCA trajectories are globally quasi-convergent (i.e. $\lim_{t\to\infty} \dot{\mathbf{a}}(t) = 0$).
- The previous line of reasoning also lets us conclude that

$$-\dot{\mathbf{u}} = \mathbf{u} - \mathbf{a} + \Phi^T \Phi \mathbf{a} - \Phi^T \mathbf{y} \in \partial V(\mathbf{a}).$$

- By the generalized chain rule (9) and (4),

$$\frac{dV}{dt} = -\dot{\mathbf{u}}(t)^T \dot{\mathbf{a}}(t) = -\sum_{n\in\Gamma} f'(u_n(t))|\dot{u}_n(t)|^2. \qquad (13)$$

- (13) is nonpositive for all $t \geq 0$ and negative for all $t \geq 0$ such that $||\dot{\mathbf{u}}_\Gamma||_2 \neq 0$.

Part (b) proof sketch

- Since $V(\mathbf{a})$ is continuous, nonincreasing, and bounded below by 0, $V(\mathbf{a}) \to V^*$, so

$$\lim_{t\to\infty} \dot{V}(\mathbf{a}) = 0 \tag{14}$$

- Since $f' > 0$, by (14) and (13), we conclude

$$\lim_{t\to\infty} ||\dot{\mathbf{u}}(t)||_2 = 0,$$

so by (4)

$$\lim_{t\to\infty} ||\dot{\mathbf{a}}(t)||_2 = 0,$$

proving global quasi-convergence of $\mathbf{a}(t)$.

Part (c) proof sketch

- Want to show: if critical points of $V(\mathbf{a})$ are isolated, LCA $\mathbf{u}(t)$ converges to a global fixed point.
- Global quasi-convergence of $\mathbf{a}(t)$ and isolated $V(\mathbf{a})$ critical points together imply convergence of (LCA) to an isolated critical point $\mathbf{a}^*$ of $V(\mathbf{a})$.
- Define $\tilde{\mathbf{a}}(t) = \mathbf{a}(t) - \mathbf{a}^*$ and $\mathbf{u}^* = -\Phi^T\Phi\mathbf{a}^* + \Phi^T\mathbf{y} + \mathbf{a}^*$, then rewrite $\dot{\mathbf{u}}(t)$ as

$$\dot{\mathbf{u}}(t) = -\mathbf{u}(t) + \mathbf{u}^* - (\Phi^T\Phi - I)\tilde{\mathbf{a}}(t) \tag{15}$$

Part (c) proof sketch

- The linear inhomogeneous ODE (15) can be solved by variation of parameters to yield

$$\mathbf{u}(t) = \mathbf{u}^* + e^{-t}(\mathbf{u}(0) - \mathbf{u}^*) + e^{-t}\int_0^t e^s(\Phi^T\Phi - I)\tilde{\mathbf{a}}(s)\,ds \quad (16)$$

- To show $\mathbf{u}(t) \to \mathbf{u}^*$, compare the expression in (16) with the linear trajectory $\boldsymbol{\ell}(t) = \mathbf{u}^* + e^{-t}(\mathbf{u}(0) - \mathbf{u}^*)$.
- Define $\mathbf{h}(t) \triangleq \mathbf{u}(t) - \boldsymbol{\ell}(t) = e^{-t}\int_0^t e^s(\Phi^T\Phi - I)\tilde{\mathbf{a}}(s)\,ds$. Global convergence of $\mathbf{u}(t)$ is proven by showing

$$\lim_{t\to\infty} ||\,\mathbf{h}(t)||_2 \le \lim_{t\to\infty} e^{-t}||\Phi^T\Phi - I||_2 \int_0^t e^s||\tilde{\mathbf{a}}(s)||_2\,ds = 0.$$

Theorem 2

- Statement: If (LCA) converges to a fixed point $\mathbf{u}^*$ such that $|u_n^*| \neq \lambda$ for all n, the system converges after a finite number of switches.
- Proof sketch:
  - Convergence of $\mathbf{u}(t)$ to $\mathbf{u}^*$ and equivalence of L^p norms on $\mathbb{R}^N$ mean that for some $t_\varrho < \infty$, $||\,\mathbf{u}(t) - \mathbf{u}^*\,||_\infty < \varrho \triangleq \min_n ||u_n^*| - \lambda|$ for all $t > t_\varrho$.
  - All such $\mathbf{u}(t)$ have the same active set as $\mathbf{u}^*$ and cannot switch active sets without leaving the set $||\,\mathbf{u}(t) - \mathbf{u}^*\,||_\infty < \varrho$.

Theorem 3 - background

- The dynamical system in (7) is exponentially convergent with convergence speed $c > 0$ if for any initial point $\mathbf{x}(0)$, there exists a constant $\kappa_0 > 0$ such that

$$||\,\mathbf{x}(t) - \mathbf{x}^*\,|| \leq \kappa_0 e^{-ct}, \quad \forall t \geq 0.$$

- Important quantity - α, bounding the magnitude of f' over the course of a trajectory.

$$\forall t \geq 0, \forall n = 1, \ldots, N \qquad |f'(u_n(t))| \leq \alpha. \tag{17}$$

Theorem 3 - background

- Important quantity - δ satisfying RIP-like condition on a subset of $\mathbb{R}^N$:

$$(1-\delta)||\,\mathbf{x}\,||_2^2 \leq ||\Phi\,\mathbf{x}\,||_2^2 \leq (1+\delta)||\,\mathbf{x}\,||_2^2$$

  for $\mathbf{x} \in \mathbb{R}^N$ with active set $\tilde{\Gamma}$.

- Given an LCA trajectory $\mathbf{u}(t)$ converging to $\mathbf{u}^*$ with active set Γ_*, $\tilde{\Gamma}$ is the union of Γ_* and every set Γ which is an active set of $\mathbf{u}(t)$ for some $t \geq 0$.

- Recall the time constant τ from (LCA). WLOG $\tau = 1$ for the proofs of Theorems 1 and 2, but it's important here to quantify convergence rates.

Theorem 3 - result

- Statement: Given an admissible thresholding function T_λ, if the previously-defined constants α and δ satisfy

$$\delta < \min\left(1, \frac{1}{(2\alpha - 1)^2}\right),$$

the LCA system is globally exponentially convergent to a unique equilibrium, with convergence speed $(1-\delta)/\tau$.

Theorem 3 - proof sketch

- Given an LCA trajectory $\mathbf{u}(t)$ converging any fixed point $\mathbf{u}^*$, define $\tilde{\mathbf{a}}(t) = \mathbf{a}(t) - \mathbf{a}^*$, $\tilde{\mathbf{u}}(t) = \mathbf{u}(t) - \mathbf{u}^*$.
- The behavior of $\frac{1}{2}||\tilde{\mathbf{u}}(t)||_2^2$ is controlled by the behavior of

$$E(t) = \frac{1}{2}||\tilde{\mathbf{u}}(t)||_2^2 + \sum_{n=1}^{N} \int_0^{\tilde{u}_n(t)} g_n(s)\, ds,$$

  where $g_n(s) = T_\lambda(s + u_n^*) - T_\lambda(u_n^*)$.
- Compare: $\tilde{a}_n(t) = T_\lambda(\tilde{u}_n(t) + u_n^*) - T_\lambda(u_n^*)$.

Theorem 3 - proof sketch

- Using the result of a technical lemma, it is demonstrated that

$$\tau \dot{E}(t) \leq -(1-\delta)E(t).$$

- Since $\frac{1}{2}||\tilde{\mathbf{u}}(t)||_2^2 \leq E(t)$, use Gronwall's inequality to conclude that

$$\frac{1}{2}||\,\mathbf{u}(t) - \mathbf{u}^*\,||_2^2 \leq E(t) \leq e^{-(1-\delta)t/\tau}E(0).$$

Table of Contents

Conclusions

- We have reviewed the problem of computing sparse representations, and the LCA system as a tool for this problem.
- The LCA system demonstrates robust convergence properties which suggest it may be feasibly implementable for efficiently solving large-scale problems.
- The theory underpinning this system connects the optimization problems of computational harmonic analysis with new dynamical systems theory for Lyapunov-like functions.
- Network implementations of LCA systems suggest intriguing connections with neuroscience.

Connections to neuroscience

- Linear-nonlinear models of neural firing rates take the form $F(\mathbf{k} \cdot \mathbf{y})$ for a receptive field $\mathbf{k}$ and a nonlinear rectification function F (thresholding and saturating).
- Competitive inhibition: Spiking neurons in excitatory populations trigger inhibitory neurons which silence large portions of the excitatory population.
- Found to occur in the mushroom body in insects, hypothesized to be involved in gamma frequency activity in mammalian visual processing.
- LCA can be implemented in a network with similar but not identical structure.

Terrible visualization

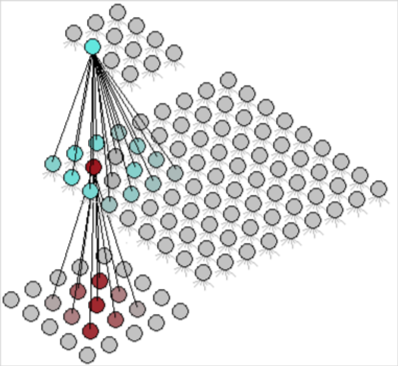


Figure: Top layer - inhibitory neurons providing feedback inhibition. Middle layer - multivalent neurons representing $\{\varphi_n\}$. Bottom layer - Input domain.

Future work

- Spiking models - what can be said about the activity of spiking neurons, whose average firing rates obey LCA-like equations?
- Positivity restrictions - can a similar system with univalent neuron firing rates be devised?
- Saturation - what sort of representations are formed when T_λ saturates as well as thresholds?
- Dictionary learning - Able to be integrated into the prior network visualization by varying bottom-middle connection strengths.

Thanks!

Works cited I

Aurele Balavoine, Justin Romberg, and Christopher J Rozell.

Convergence and rate analysis of neural networks for sparse approximation.

Neural Networks and Learning Systems, IEEE Transactions on, 23(9):1377–1389, 2012.

Adam S Charles, Pierre Garrigues, and Christopher J Rozell.

A common network architecture efficiently implements a variety of sparsity-based inference problems.

Neural computation, 24(12):3317–3339, 2012.

Works cited II

John J Hopfield.

Neural networks and physical systems with emergent collective computational abilities.

Proceedings of the national academy of sciences, 79(8):2554–2558, 1982.

Andrew C Lin, Alexei M Bygrave, Alix de Calignon, Tzumin Lee, and Gero Miesenböck.

Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination.

Nature neuroscience, 2014.

Works cited III

Bruno A Olshausen and David J Field.

Sparse coding of sensory inputs.

Current opinion in neurobiology, 14(4):481–487, 2004.

Christopher J Rozell, Don H Johnson, Richard G Baraniuk, and Bruno A Olshausen.

Sparse coding via thresholding and local competition in neural circuits.

Neural computation, 20(10):2526–2563, 2008.