# Certain Connections between Convex Quadratic Optimization by Artificial Analog Neural Networks and Traditional Methods of Mathematical Programming


Ilan Adler
University of California at Berkeley

Sushil Verma
University of Southern California


October 24, 1995

## 1 Neural Network Model

The neural network model we use in this paper is the continuous-valued version of the McCulloch and Pitts model whose dynamics is given by the following equation:

$$\dot{u} = -u + d + Wg(u) \tag{1.1}$$

where $g : \mathcal{Q} \to \mathcal{R} \quad (\mathcal{Q} \subset \mathcal{R}), \quad \tau \in \mathcal{R}_{++}, \quad u \in \mathcal{Q}^n, \quad d \in \mathcal{R}^n, \quad W \in \mathcal{R}^{n \times n}$. Suppose $\mathcal{Q}$ is an open interval in $\mathcal{R}$ and $q \in \mathcal{Q}$. We call $g$ an *activation function.* Various researchers ([9], [2], [6], [14]) have analyzed this dynamics in detail. Convergence and basin of convergence are two main issues of interest.

Of special interest are the neural networks which have a symmetric weight matrix $W$. The dynamics of the above neural networks have been known to be such that it tends to the minima of a certain quadratic function. This property is at the basis of their use as associative memory devices. On the other hand, various researchers in the field of electrical engineering have realized and explored the possibility of using these neural networks as devices to perform quadratic optimization. These reseachers often implement the neural networks as analog electrical circuits, and borrow freely from the theory of electrical circuits to analyse the properties of those circuits. What has been rarely seen is a treatment of the topic based on the tremendous amount of work which has been done in the area of optimization of quadratic function in the field of mathematical programming. The aim of this paper is primarily to fill that gap and raise some interesting questions in the process.

## 2 Quadratic Optimization Problem

A fundamental problem in convex quadratic optimization is the following.

**Problem 2.1** Convex Quadratic Optimization

| | Primal | | | Dual | |
|---|---|---|---|---|---|
| min | $1/2x^tQx + c^tx$ | | min | $1/2v^TQv$ | |
| subject to | $x$ | $\geq 0$ | subject to | $-Qv$ | $\leq c$ |

Here, $Q$ is a symmetric positive semi-definite matrix. If we denote the slacks as $z := c + Qv$ and denote the pseudo-inverse of the matrix $Q$ as $Q^+$, then the dual of the above problem can be written in a different form.

**Problem 2.2** Convex Quadratic Optimization - Alternate Form

| | Primal | | | | Dual | |
|---|---|---|---|---|---|---|
| min | $1/2x^tQx + c^tx$ | | | min | $1/2(z-c)^tQ^+(z-c)$ | |
| subject to | $x$ | $\geq 0$ | | subject to | $[I - QQ^+](z-c)$ | $= 0$ |
| | | | | | $z$ | $\geq 0$ |

In this paper we will mostly deal with the alternate form of the convex quadratic optimization problem. How can this primal-dual pair of quadratic problems be mapped to the single neural dynamics? The following section sets up the framework to answer the above question.

# 3 Convergence

To analyze the convergence properties of the dynamics we will use the Generalized Liapunov Result stated below. For a proof, the reader should refer to [5].

Let $x = x(t)$ be a solution of the following autonomous dynamical system.

$$\dot{x} = f(x) \tag{3.2}$$

where the function $f$ is continuously differentiable in the domain of interest.

**Definition 3.1** A set $G$ is an *invariant set* for dynamic system 3.2 if whenever a point $x$ on a system trajectory is in $G$, the trajectory remains in $G$.

**Theorem 3.2 Generalized Liapunov Result - Invariant Set Theorem**
*Suppose that*

**(a)** $V(x) \in C^2$ *is a scalar function.*

**(b)** *The set* $\Omega_s = \{x : V(x) \leq s\}$ *is bounded.*

**(c)** $\dot{V}(x) \leq 0$ *within* $\Omega_s$.

**(d)** $S$ *is the set of points within* $\Omega_s$ *where* $\dot{V}(x) = 0$, *and* $G$ *is the largest invariant set within* $S$.

*Then every trajectory in* $\Omega_s$ *tends to* $G$ *as time increases.* ∎

The scalar function, $V$, in the above theorem is often called an *energy function* or a *potential function.* In mathematics literature, it is also often called a *Liapunov function.*

Existence of a energy function for a dynamic system is often an indication of its good behavior. The energy function helps in the analysis of the dynamic system because it is a scalar function that embodies a trajectory in multi-dimensional space.

We present **two** different functions, each of which acts as an energy function for the dynamics 1.1 under certain conditions in certain appropriate domain. Later, we will draw a parallel between this energy function and certain other functions which are to be optimized.

We label the system under the first and second energy functions as *Primal Model* and *Dual Model* respectively. The reason to call these models primal and dual will become clear later in the paper.

### 3.1 Primal Model

The associated energy function is:

$$f_p(\boldsymbol{u}) = -\tfrac{1}{2}g(\boldsymbol{u})^T\boldsymbol{W}g(\boldsymbol{u}) - \boldsymbol{d}^Tg(\boldsymbol{u}) + \sum_{j=1}^{n}\int_q^{u_j} g^{'}(v)vdv \tag{3.3}$$

This energy function was first presented in this context by Hopfield and Tank [10]. Since then it has been the basis of various other algorithms in the field of neural networks.

The following two propositions use Theorem 3.2 to prove a convergence result in our context.

**Proposition 3.3** *Suppose that*

**(i)** $g(\mathcal{Q} \to \mathcal{R}) \in \mathcal{C}^2$.

**(ii)** $g^{'} > 0$ *(i.e. g is strictly increasing).*

**(iii)** $\boldsymbol{W}$ *is symmetric.*

**(iv)** *Starting at* $\boldsymbol{u}^o \equiv \boldsymbol{u}(t_0), \boldsymbol{u}(t) \in \mathcal{Q}^n$ *for* $t \geq t_0$.

**(v)** *The set* $\Omega_p = \{\boldsymbol{u} \in \mathcal{Q}^n : f_p(\boldsymbol{u}) \leq f_p(\boldsymbol{u}(t_0))\}$ *is bounded.*

*then, any trajectory induced by 1.1 and starting at* $\boldsymbol{u}^0$ *tends to the set* $\{\boldsymbol{u} : \dot{\boldsymbol{u}} = \boldsymbol{0}\}$. *Moreover, the trajectory converges to a first-order minima of* $f_p$.

**Proof :** Since $g \in \mathcal{C}^2$ by (i), $f_p \in \mathcal{C}^2$. We now prove that the energy function strictly goes down with time.

$$\nabla f_p(\boldsymbol{u}) = (-\boldsymbol{W}g(\boldsymbol{u}) - \boldsymbol{d} + \boldsymbol{u})g^{'}(\boldsymbol{u}) = -\dot{\boldsymbol{u}}g^{'}(\boldsymbol{u})$$
$$\dot{f}_p = \dot{\boldsymbol{u}}^T\nabla f_p(\boldsymbol{u}) = -\dot{\boldsymbol{u}}^T(\dot{\boldsymbol{u}}g^{'}(\boldsymbol{u})) \leq 0$$

We used the fact that $\boldsymbol{W}$ is symmetric for the derivation in the first equation and the fact that $g$ is a strictly increasing function for the second equation.

If we let $f_p$ be the scalar function in Theorem 3.2 then requirements (a) and (c) are satisfied. Moreover, requirement (b) is implied by assumptions (iv) and (v) directly. Hence, the function $f_p$ is a Liapunov function for the dynamics 1.1 and using the theorem, we can conclude that any trajectory induced by 1.1 and starting at $\boldsymbol{u}^0$ tends to the set $\{\boldsymbol{u} : \dot{f}_p(\boldsymbol{u}) = \boldsymbol{0}\}$. But note that $\dot{f}_p = 0 \Leftrightarrow \dot{\boldsymbol{u}} = \boldsymbol{0}$. Thus the first part of the result follows. To see the second part, note that $\nabla f_p(\boldsymbol{u}) = -\dot{\boldsymbol{u}}g^{'}(\boldsymbol{u}) = 0$. ∎

### 3.2 Dual Model

The associated energy function is:

$$f_d(\boldsymbol{u}) = -\tfrac{1}{2}(\boldsymbol{u} - \boldsymbol{d})^T\boldsymbol{W}^+(\boldsymbol{u} - \boldsymbol{d}) + \sum_{j=1}^{n}\int_q^{u_j} g(v)dv \tag{3.4}$$

where, $\boldsymbol{W}^+$ is the Penrose-Moore pseudo-inverse of the matrix $\boldsymbol{W}$. For definition of the Penrose-Moore pseudo-inverse, see [12].

This energy function is interesting in the way it operates. It is not valid in the entire domain as the energy function associated with primal model. Instead, it is operative on a certain affine space. It differs in two more ways. It needs $-\boldsymbol{W}$ to be positive semi-definite and the activation function need not be monotonic.

The following proposition provides the details:

**Proposition 3.4** *Suppose that*

(i) $g(\mathcal{Q} \rightarrow \mathcal{R}) \in \mathcal{C}^2$.

(ii) $-\boldsymbol{W}$ *is symmetric and positive semi-definite.*

(iii) *Starting at* $\boldsymbol{u}^o \equiv \boldsymbol{u}(t_0), \boldsymbol{u}(t) \in \mathcal{Q}^n$ *for* $t \geq t_0$.

(iv) $\boldsymbol{P}(-\boldsymbol{u}^0 + \boldsymbol{d}) = \boldsymbol{0}$ *(where* $\boldsymbol{P} \equiv \boldsymbol{I} - \boldsymbol{W}^+\boldsymbol{W}$*).*

(v) *The set* $\Omega_d = \{\boldsymbol{u} \in \mathcal{Q}^n : \boldsymbol{P}\boldsymbol{u} = \boldsymbol{P}\boldsymbol{d}, \quad f_d(\boldsymbol{u}) \leq f_d(\boldsymbol{u}^0))\}$ *is bounded.*

*then, any trajectory induced by 1.1 and starting at* $\boldsymbol{u}^0$ *tends to the set* $\{\boldsymbol{u} : \dot{\boldsymbol{u}} = \boldsymbol{0}\}$. *Moreoever, the trajectory converges to a first order minima of* $f_d(\boldsymbol{u})$ *restricted to the affine space* $\boldsymbol{P}(-\boldsymbol{u} + \boldsymbol{d}) = 0$.

**Proof :** ¿From the definition of the Penrose-Moore pseudo-inverse, $\boldsymbol{W}^+\boldsymbol{W}\boldsymbol{W}^+ = \boldsymbol{W}^+$, $\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{W} = \boldsymbol{W}$, $\boldsymbol{W}\boldsymbol{W}^+ = (\boldsymbol{W}\boldsymbol{W}^+)^T$ and $\boldsymbol{W}^+\boldsymbol{W} = (\boldsymbol{W}^+\boldsymbol{W})^T$. These properties and the given fact that $-\boldsymbol{W}$ is symmetric positive semi-definite implies that $-\boldsymbol{W}^+$ is also symmetric positive semi-definite. (See [8] for proofs.) That in turn implies $\boldsymbol{W}^+\boldsymbol{W} = \boldsymbol{W}\boldsymbol{W}^+$. Moreover, $\boldsymbol{P}\boldsymbol{W} = \boldsymbol{P}\boldsymbol{W}^+ = 0$.

It is easy to show that all the trajectories which start on the affine space $\boldsymbol{P}\boldsymbol{u} = \boldsymbol{P}\boldsymbol{d}$ stay on it. This follows from the calculation at $\boldsymbol{u} = \boldsymbol{u}^0$.

$$\begin{aligned}
\tau \boldsymbol{P}\dot{\boldsymbol{u}} &= \boldsymbol{P}(-\boldsymbol{u}^0 + \boldsymbol{d}) + \boldsymbol{P}\boldsymbol{W}\boldsymbol{g}(\boldsymbol{u}) \\
&= \boldsymbol{P}(-\boldsymbol{u}^0 + \boldsymbol{d}) \\
&= 0 (\text{using assumption (iv)})
\end{aligned}$$

All the higher derivatives are similarly in the null space of $\boldsymbol{P}$ at $\boldsymbol{u} = \boldsymbol{u}^0$. Hence, using the Taylor series expansion, $\boldsymbol{P}\dot{\boldsymbol{u}}$ is zero everywhere on the mentioned affine space.

Since $\boldsymbol{u}(t)$ stays on the affine space, we need to look at the derivative of the energy function restricted to this affine space only.

$$\begin{aligned}
\nabla f_d(\boldsymbol{u}) &= -\boldsymbol{W}^+(\boldsymbol{u} - \boldsymbol{d}) + \boldsymbol{g}(\boldsymbol{u}) \\
&= -\boldsymbol{W}^+(-\dot{\boldsymbol{u}} + \boldsymbol{W}\boldsymbol{g}(\boldsymbol{u})) + \boldsymbol{g}(\boldsymbol{u}) \\
&= +\boldsymbol{W}^+\dot{\boldsymbol{u}} + \boldsymbol{P}\boldsymbol{g}(\boldsymbol{u})
\end{aligned}$$

$$\begin{aligned}
\dot{f}_d(\boldsymbol{u}) &= \dot{\boldsymbol{u}}^T \nabla f_d(\boldsymbol{u}) \\
&= +\dot{\boldsymbol{u}}^T \boldsymbol{W}^+ \dot{\boldsymbol{u}} + \dot{\boldsymbol{u}}^T \boldsymbol{P}\boldsymbol{g}(\boldsymbol{u}) \\
&= +\dot{\boldsymbol{u}}^T \boldsymbol{W}^+ \dot{\boldsymbol{u}} + 0 \\
&\leq 0
\end{aligned}$$

In fact, $\dot{f}_d(\boldsymbol{u})$ equals zero if and only if $\dot{\boldsymbol{u}}$ is zero. Suppose that $\dot{\boldsymbol{u}}\boldsymbol{W}^+\dot{\boldsymbol{u}} = \boldsymbol{0}$. Since, $-\boldsymbol{W}^+$ is positive semi-definite, this implies that $\boldsymbol{W}^+\dot{\boldsymbol{u}} = 0$. Since $\boldsymbol{P}\dot{\boldsymbol{u}} = \dot{\boldsymbol{u}} - \boldsymbol{W}^+\boldsymbol{W}\dot{\boldsymbol{u}} = 0$, $\dot{\boldsymbol{u}} = \boldsymbol{W}^+\boldsymbol{W}\dot{\boldsymbol{u}} = \boldsymbol{W}\boldsymbol{W}^+\dot{\boldsymbol{u}} = 0$. Hence, $\dot{f}_d(\boldsymbol{u}) = 0 \Leftrightarrow \dot{\boldsymbol{u}} = \boldsymbol{0}$.

Also, since $g \in \mathcal{C}^2$ by (i), $f_d \in \mathcal{C}^2$. If we let $f_d$ be the scalar function in Theorem 3.2 then the requirements (a) and (c) are satisfied. Moreover, the requirement (b) is implied by assumptions (iii)-(v) directly. Hence, the function $f_d$ is a *Liapunov function* for the dynamics 1.1, which implies that any trajectory induced by 1.1 and starting at $u^0$ tends to the set $\{u : \dot{f}_d(u) = 0\}$. But as we proved above, $\dot{f}_d(u) = 0 \Leftrightarrow \dot{u} = 0$. Hence the first part of the result follows. For the second part, recall that the first order conditions for $u'$ to be a minima of $f_d(u)$ subject to $P(-u + d) = 0$ is that $\nabla f_d(u') + Py = 0$ for some vector $y$. Now note that $\nabla f_d(u) = Pg(u)$ in limit as $\dot{u} = 0$. Therefore, in limit the first-order minima conditions are satisfied if we substitute $y = g(u)$. ■

The conditions required for convergence under the two models are different. The following corollary collects all the common conditions.

**Corollary 3.5** *Suppose that*

**(i)** $g(\mathcal{Q} \rightarrow \mathcal{R}) \in \mathcal{C}^2$.

**(ii)** $g' > 0$ *(i.e. g is strictly increasing).*

**(iii)** $-W$ *is symmetric and positive semi-definite.*

**(iv)** *Starting at* $u^o \equiv u(t_0), u(t) \in \mathcal{Q}^n$ *for* $t \geq t_0$.

**(v)** $P(-u^0 + d) = 0$ *(where* $P \equiv I - W^+W$*).*

**(vi)** *The set* $\Omega_p = \{u \in \mathcal{Q}^n : f_p(u) \leq f_p(u(t_0))\}$ *is bounded.*

**(vii)** *The set* $\Omega_d = \{u \in \mathcal{Q}^n : Pu = Pd, \quad f_d(u) \leq f_d(u^0))\}$ *is bounded.*

*then, any trajectory induced by 1.1 and starting at* $u^0$ *tends to the set* $\{u : \dot{u} = 0\}$. *Moreover, the trajectory converges to a first order minima of* $f_d$ *(subject to the affine constraints* $P(u + d) = 0$*) and* $f_p$.

**Proof :** Follows directly from Proposition 3.3 and 3.4. ■

Here, the reader should note that the conditions under which the two energy functions operate are different. However, even under conditions which satisfy the requirements for both, the two energy functions are still not identical. Therefore, the difference in two energy functions is not just on the surface.

A question which arises is what is the physical significance of two different energy functions for the same neural dynamics. The answer lies in the physical implementation of the above dynamics. There are two natural ways of implementing this dynamics, one where the primary variable is $u$ and the other where the primary variable is $g(u)$. Depending upon which one is the primary variable, a physical system can have two different kind of potential functions with some physical significance. Indeed, the above dynamics can be implemented using electrical circuits in two different ways. One method uses non-linear resistors, while the other uses non-linear capacitors.

There is obvious similarity between primal model energy function and the primal objective function of Problem 2.2 (with the substitution $x = g(u)$ and again between dual model energy function and the dual objective function of Problem 2.2 (with the substitution $z = -u$). The only difference is a term solely involving activation functions. In the next two sections, we look at these similarities more closely. We study the role of the extra term which involves activation functions.

# 4 Optimization

We present three different interpretations of the neural dynamics in this section. We use Proposition 3.3 for a primal interpretation, Proposition 3.4 for a dual interpretation and Corollary 3.5 for a primal-dual interpretation.

Throughout this section we use the following specializations.

$$W = -Q \tag{4.5}$$
$$d = -c \tag{4.6}$$

We will denote the inverse function $g^{-1}$ by $h$ if it is well-defined. We further denote $g(u)$ by $x$ and $u$ by $-z$. Under these new notations,

$$f_p(x) = 1/2x^TQx + c^Tx + \sum_j \int_{g(q)}^{x_j} h(v)dv, \tag{4.7}$$
$$f_d(z) = 1/2(z-c)^TQ^+(z-c) - \sum_j \int_{-q}^{z_j} g(-v)dv. \tag{4.8}$$

Note that functions $f_p$ and $f_d$ minus their last term involving the activation function are identical to the primal and dual objective function respectively of Problem 2.2. Denote the last term of $f_p$ and $f_d$ as $r_p$ and $r_d$ respectively.

$$r_p(x) = \sum_j \int_{g(q)}^{x_j} h(v)dv \tag{4.9}$$
$$r_d(z) = -\sum_j \int_{-q}^{z_j} g(-v)dv \tag{4.10}$$

Suppose $\{g_\beta : \beta > 0\}$ is a family of activation functions smoothly parameterized by a single parameter $\beta$. Denote the corresponding last terms as $r_p^\beta$ and $r_d^\beta$.

## 4.1 Primal Trajectories

**Proposition 4.1** *Suppose that*

**(i)** $g_\beta : Q_\beta \rightarrow (-\alpha_\beta, \infty)$ *for some* $\alpha_\beta > 0$,

**(ii)** $g_\beta \in C^2$ *is an onto strictly increasing function,*

**(iii)** *For* $v > 0$, $\lim_{\beta\rightarrow 0} r_p^\beta(v) =$ *constant,*

**(iv)** *For* $v < 0$, $\lim_{\beta\rightarrow 0} r_p^\beta(v) = \infty$,

**(v)** *Either* $\alpha_\beta = \infty$, *or* $\lim_{v\rightarrow -\alpha_\beta} h_\beta(v) = \infty$.

**(v)** *The primal level sets of Problem 2.2 are bounded, and*

**(vi)** *There is available a vector* $x^0$ *such that* $x^0 \in (-\alpha_\beta, \infty)^n$.

*Then, starting at* $u^0 = h(x^0)$ *the neural network whose dynamic is given in 1.1 converges to* $x^\beta$. *Moreover, for a small enough value of* $\beta$, $x^\beta$ *is an approximate primal solution to Problem 2.2 to the required degree.*

### 4.2 Dual Trajectories

**Proposition 4.2** *Suppose that*

**(i)** $g_\beta : (-\infty, \gamma_\beta) \to \mathcal{R}$ *for some* $\gamma_\beta > 0$,

**(ii)** $g_\beta \in \mathcal{C}^2$,

**(iii)** *For* $v > 0$, $\lim_{\beta \to 0} \; r_d^\beta(v) = constant$,

**(iv)** *For* $v < 0$, $\lim_{\beta \to 0} \; r_d^\beta(v) = \infty$,

**(v)** *Either* $\gamma_\beta = \infty$, *or* $\lim_{v \to -\gamma_\beta} g_\beta(v) = \infty$.

**(v)** *The dual level sets of Problem 2.2 are bounded, and*

**(vi)** *There is available a vector* $\boldsymbol{z}^0$ *such that* $(\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^+)(\boldsymbol{z}^0 - \mathbf{c}) = 0$ *and* $\boldsymbol{z}^0 \in (-\gamma_\beta, \infty)^n$.

*Then, starting at* $\boldsymbol{u} = -\boldsymbol{z}^0$ *the neural network, whose dynamic is given in 1.1, converges generating* $\boldsymbol{z}^\beta$ *satisfying* $(\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^+)(\boldsymbol{z}^\beta - \mathbf{c}) = 0$. *Moreover, for a small enough value of* $\beta$, $\boldsymbol{z}^\beta$ *is an approximate dual solution to Problem 2.2 to the required degree.*

### 4.3 Primal-Dual Trajectories

**Proposition 4.3** *Suppose that*

**(i)** $g_\beta : (-\infty, \gamma_\beta) \to (-\alpha_\beta, \infty)$ *for some* $\alpha_\beta, \gamma_\beta > 0$

**(ii)** $g_\beta \in \mathcal{C}^2$ *is an onto strictly increasing function,*

**(iii)** *For* $v > 0$, $\lim_{\beta \to 0} \; r_p^\beta(v) = constant$ *and* $\lim_{\beta \to 0} \; r_d^\beta(v) = constant$,

**(iv)** *For* $v < 0$, $\lim_{\beta \to 0} \; r_p^\beta(v) = \infty$ *and* $\lim_{\beta \to 0} \; r_d^\beta(v) = \infty$,

**(v)** *The primal and dual level sets of Problem 2.2 are bounded, and*

**(vi)** *There is available a vector* $\boldsymbol{z}^0$ *such that* $(\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^+)(\boldsymbol{z}^0 - \mathbf{c}) = 0$ *and* $\boldsymbol{z}^0 \in (-\alpha_\beta, \infty)^n$.

*Then, starting at* $\boldsymbol{u} = -\boldsymbol{z}^0$ *the neural network, whose dynamic is given in 1.1, converges generating* $\boldsymbol{z}^\beta$ *satisfying* $(\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^+)(\boldsymbol{z}^\beta - \mathbf{c}) = 0$. *Moreover, for a small enough value of* $\beta$, $\boldsymbol{x}^\beta = \boldsymbol{g}(\boldsymbol{u})$ *and* $\boldsymbol{z}^\beta$ *are approximate primal and dual solution respectively to Problem 2.2 to the required degree.*

## 5 Relationship to Known Continuous Trajectories

The natural question to ask at this stage is: what relationship, if any, does the previous trajectories have to previously known continuous trajectories in the mathematical programming literature.

### 5.1 Primal

In terms of primal variables $x$ the neural trajectories can be written down in the following form.

$$\dot{x} = -\frac{1}{h'_\beta(x)}(Qx + c + h_\beta(x)) \tag{5.11}$$

It is apparent that the trajectory $\{x(t) : t > 0\}$ is not a gradient descent for any scalar function. However, one can show that it follows gradient of the energy function $f_p$ under a non-Euclidean Riemannian metric. It is known that any $n \times n$ symmetric and positive definite matrix $M$ whose entries $m_{ij}$ are smooth function of the coordinates defines a Riemannian Metric. We can use $M = \text{diag}(h'_\beta(x_1), h'_\beta(x_2), ...., h'_\beta(x_n))$. It can be shown that gradient of a function $f$ with respect to metric $M$, is $M^{-1}\nabla f$. Therefore, the gradient of the function $f_p$ with respect to metric $M$, is exactly the negative of the right hand side of the dynamics 5.11.

To see another link, consider the dynamics 5.11 with the specific function $h_\beta(x) = -\beta/x$.

$$\beta\dot{x} = -X^2(Qx + c - \beta X^{-1}e) \tag{5.12}$$

The notation $X$ denotes a diagonal matrix with $X_{ii} = x_i$ and $e$ denotes a vector of 1's. The trajectory $\{x(t) : t > 0\}$ is identical (up to a scaling factor $\beta$) to the so called $\mu$-Barrier Continuous Trajectory described in [13] applied to Problem 2.2. It has been studied and compared to the previous two trajectories by Gill et al. in [7] and Shub in [15].

These trajectories are closely related to affine-scaling trajectories in the sense that for $\beta = 0$, they are identical to them. For a more detailed presentation, see [1] and a series of papers [3], [4] and [11].

### 5.2 Dual

In terms of dual variables $z$ the neural trajectories can be written down in the following form.

$$\dot{z} = -z + c + Q^+g(-z) \tag{5.13}$$

Recall that the trajectory $\{z(t) : t > 0\}$ stays on the affine space $(I - QQ^+)(z - c) = 0$. It is not clear if this trajectory follows the gradient of any scalar function.

Specifically, consider the energy function $f_d$. The gradient of $f_d$ is $Q^+(z - c) - g(-z)$. This gradient projected on the affine space $(I - QQ^+)(z - c) = 0$ is $QQ^+(Q^+(z - c) - g(-z))$ Using the fact that $QQ^+ = Q^+Q$, we get the projected gradient as $z - c - QQ^+g(-z)$. On comparing the right hand side of equation 5.13, one can conclude that the above trajectory is a projected steepest descent trajectory for function $f_d$ if and only if $QQ^+ = Q^+$ or equivalently $Q^+(= Q)$ is a projection matrix.

### 5.3 Primal-Dual

The primal-dual trajectory $\{(x(t) = g(u(t)), z(t) = -u(t)) : t > 0\}$ is such that $\{x(t) : t > 0\}$ and $\{z(t) : t > 0\}$ have the properties of a primal and dual trajectory respectively, which are described in the previous two sections.

Under the special case when $Q^+$ is a projection matrix, an interesting conclusion can be drawn. If the trajectory $z(t)$ starts on the mentioned affine space, then it is a projected steepest descent trajectory for function $f_d$. On the other hand, trajectory $x(t) = g(-z(t))$ is a $\mu$-Barrier trajectory which is also a a steepest descent trajectory with respect to a well-defined Riemannian metric. Therefore, the properties associated with $\mu$-Barrier trajectories can be associated with projected steepest descent trajectories under these conditions.

## References


[1] I. Adler and R. Monteiro. Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Contemporary Mathematics*, 114:189–211, 1990.

[2] A. Atiya. Learning on a general network. *Proceedings of IEEE Conference on Neural Information Processing Systems at Denver, Colorado*, November 1987.

[3] D.A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. i: Affine an projective scaling trajectories. *Transactions of the American Mathematical Society*, 314(2):499–526, Aug 1989.

[4] D.A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. ii: Legendre transform coordinates and central trajectories. *Transactions of the American Mathematical Society*, 314(2):527–581, Aug 1989.

[5] Fred Brauer and John A. Nohel. *Qualitative Theory of Ordinary Differential Equations: An Introduction.* W.A. Benjamin, Inc. New York, 1969.

[6] M.A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 13:815–826, 1983.

[7] P.E. Gill, W. Murrray, Tomlin M. A., and Wright M. H. On projected newton barrier methods for linear programming and an equivalence to karmarkar's projection method. Technical Report SOL 85-11, Department of Operations Research, Stanford University, Stanford, CA 94305, 1985.

[8] F. A. Graybill. *Matrices with Applications in Statistics.* Wadsworth International Group, 1983.

[9] John J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of National Academy of Sciences, USA*, 81:3088–3092, May 1984.

[10] John J. Hopfield and D. W. Tank. "neural" computation of decisions in optimization problems. *Biological Cybernatics*, 52:141–152, 1985.

[11] J. C. Lagarias. The nonlinear geometry of linear programming. iii: Projective legendre transform coordinates and hilbert geometry. *Transactions of the American Mathematical Society*, 320:193–225, July 1990.

[12] Peter Lancaster. *Theory of Matrices.* Academic Press, New York, 1969.

[13] N. Megiddo and M. Shub. Boundary behavior of interior point algorithms in linear programming. Technical Report RJ-5319(54679), IBM Almaden Research Center, San Jose, CA, 1986.

[14] Fernando J. Pineda. Dynamics and architecture for neural computation. *Journal of Complexity*, 4:216–152, 1988.

[15] M. Shub. On the asymptotic behavior of the projective rescaling algorithm for linear programming. *Journal of Complexity*, 3:258–269, 1987.