Asymptotic Theory for Regularization: One-Dimensional Linear Case

Petri Koistinen
Rolf Nevanlinna Institute, P.O. Box 4, FIN-00014 University of Helsinki, Finland. Email: Petri.Koistinen@rni.helsinki.fi

Abstract

The generalization ability of a neural network can sometimes be improved dramatically by regularization. To analyze the improvement one needs more refined results than the asymptotic distribution of the weight vector. Here we study the simple case of one-dimensional linear regression under quadratic regularization, i.e., ridge regression. We study the random design, misspecified case, where we derive expansions for the optimal regularization parameter and the ensuing improvement. It is possible to construct examples where it is best to use no regularization.

1 INTRODUCTION

Suppose that we have available training data $(X_1, Y_1), \ldots, (X_n, Y_n)$ consisting of pairs of vectors, and we try to predict Y_i on the basis of X_i with a neural network with weight vector w. One popular way of selecting w is by the criterion

$$\frac{1}{n}\sum_{1}^{n} \ell(X_i, Y_i, w) + \lambda Q(w) = \min!, \tag{1}$$

where the loss $\ell(x, y, w)$ is, e.g., the squared error $\|y - g(x, w)\|^2$, the function $g(\cdot, w)$ is the input/output function of the neural network, the penalty $Q(w)$ is a real function which takes on small values when the mapping $g(\cdot, w)$ is smooth and high values when it changes rapidly, and the regularization parameter λ is a nonnegative scalar (which might depend on the training sample). We refer to the setup (1) as (training with) regularization, and to the same setup with the choice $\lambda = 0$ as training without regularization. Regularization has been found to be very effective for improving the generalization ability of a neural network especially when the sample size n is of the same order of magnitude as the dimensionality of the parameter vector w, see, e.g., the textbooks (Bishop, 1995; Ripley, 1996).

In this paper we deal with asymptotics in the case where the architecture of the network is fixed but the sample size grows. To fix ideas, let us assume that the training data is part of an i.i.d. (independent, identically distributed) sequence $(X,Y);(X_1,Y_1),(X_2,Y_2),\ldots$ of pairs of random vectors, i.e., for each i the pair (X_i,Y_i) has the same distribution as the pair (X,Y) and the collection of pairs is independent (X and Y can be dependent). Then we can define the (prediction) risk of a network with weights w as the expected value

$$r(w) := \mathbb{E}\,\ell(X,Y,w). \tag{2}$$

Let us denote the minimizer of (1) by $\hat{w}_n(\lambda)$, and a minimizer of the risk r by w^*. The quantity $r(\hat{w}_n(\lambda))$ is the average prediction error for data independent of the training sample. This quantity $r(\hat{w}_n(\lambda))$ is a random variable which describes the generalization performance of the network: it is bounded below by $r(w^*)$ and the more concentrated it is about $r(w^*)$, the better the performance. We will quantify this concentration by a single number, the expected value $\mathbb{E}\,r(\hat{w}_n(\lambda))$. We are interested in quantifying the gain (if any) in generalization for training with versus training without regularization defined by

$$\mathbb{E}\,r(\hat{w}_n(0)) - \mathbb{E}\,r(\hat{w}_n(\lambda)). \tag{3}$$

When regularization helps, this is positive.

However, relatively little can be said about the quantity (3) without specifying in detail how the regularization parameter is determined. We show in the next section that provided λ converges to zero sufficiently quickly (at the rate $o_p(n^{-1/2})$), then $\mathbb{E}\,r(\hat{w}_n(0))$ and $\mathbb{E}\,r(\hat{w}_n(\lambda))$ are equal to leading order. It turns out, that the optimal regularization parameter resides in this asymptotic regime. For this reason, delicate analysis is required in order to get an asymptotic approximation for (3). In this article we derive the needed asymptotic expansions only for the simplest possible case: one-dimensional linear regression where the regularization parameter is chosen independently of the training sample.

2 REGULARIZATION IN LINEAR REGRESSION

We now specialize the setup (1) to the case of linear regression and a quadratic smoothness penalty, i.e., we take $\ell(x,y,w) = [y - x^T w]^2$ and $Q(w) = w^T R w$, where now y is scalar, x and w are vectors, and R is a symmetric, positive definite matrix. It is well known (and easy to show) that then the minimizer of (1) is

$$\hat{w}_n(\lambda) = \left[\frac{1}{n}\sum_1^n X_i X_i^T + \lambda R\right]^{-1} \frac{1}{n}\sum_1^n X_i Y_i. \tag{4}$$

This is called the *generalized ridge regression estimator*, see, e.g., (Titterington, 1985); ridge regression corresponds to the choice $R = I$, see (Hoerl and Kennard, 1988) for a survey. Notice that (generalized) ridge regression is usually studied in the *fixed design* case, where X_i:s are nonrandom. Further, it is usually assumed that the model is *correctly specified*, i.e., that there exists a parameter such that $Y_i = X_i^T w^* + \epsilon_i$, and such that the distribution of the noise term ϵ_i does not depend on X_i. In contrast, we study the *random design, misspecified* case.

Assuming that $\mathbb{E}\,\|X\|^2 < \infty$ and that $\mathbb{E}[XX^T]$ is invertible, the minimizer of the risk (2) and the risk itself can be written as

$$w^* = A^{-1}\mathbb{E}[XY], \quad \text{with} \quad A := \mathbb{E}[XX^T] \tag{5}$$

$$r(w) = r(w^*) + (w - w^*)^T A (w - w^*). \tag{6}$$

If Z_n is a sequence of random variables, then the notation $Z_n = o_p(n^{-\alpha})$ means that $n^\alpha Z_n$ converges to zero in probability as $n \to \infty$. For this notation and the mathematical tools needed for the following proposition see, e.g., (Serfling, 1980, Ch. 1) or (Brockwell and Davis, 1987, Ch. 6).

Proposition 1 *Suppose that* $\mathbb{E}\,Y^4 < \infty$, $\mathbb{E}\,\|X\|^4 < \infty$ *and that* $A = \mathbb{E}\,[XX^T]$ *is invertible. If* $\lambda = o_p(n^{-1/2})$, *then both* $\sqrt{n}(\hat{w}_n(0) - w^*)$ *and* $\sqrt{n}(\hat{w}_n(\lambda) - w^*)$ *converge in distribution to* $N(0, C)$, *a normal distribution with mean zero and covariance matrix* C.

The previous proposition also generalizes to the nonlinear case (under more complicated conditions). Given this proposition, it follows (under certain additional conditions) by Taylor expansion that both $\mathbb{E}\,r(\hat{w}_n(\lambda)) - r(w^*)$ and $\mathbb{E}\,r(\hat{w}_n(0)) - r(w^*)$ admit the expansion $\beta_1 n^{-1} + o(n^{-1})$ with the same constant β_1. Hence, in the regime $\lambda = o_p(n^{-1/2})$ we need to consider higher order expansions in order to compare the performance of $\hat{w}_n(\lambda)$ and $\hat{w}_n(0)$.

3 ONE-DIMENSIONAL LINEAR REGRESSION

We now specialize the setting of the previous section to the case where x is scalar. Also, from now on, we only consider the case where the regularization parameter for given sample size n is deterministic; especially λ is not allowed to depend on the training sample. This is necessary, since coefficients in the following type of asymptotic expansions depend on the details of how the regularization parameter is determined. The deterministic case is the easiest one to analyze.

We develop asymptotic expansions for the criterion

$$J_n(k) := \mathbb{E}\,(r(\hat{w}_n(k))) - r(w^*), \tag{7}$$

where now the regularization parameter k is deterministic and nonnegative. The expansions we get turn out to be valid uniformly for $k \geq 0$. We then develop asymptotic formulas for the minimizer of J_n, and also for $J_n(0) - \inf J_n$. The last quantity can be interpreted as the average improvement in generalization performance gained by optimal level of regularization, when the regularization constant is allowed to depend on n but not on the training sample.

From now on we take $Q(w) = w^2$ and assume that $A = \mathbb{E}\,X^2 = 1$ (which could be arranged by a linear change of variables). Referring back to formulas in the previous section, we see that

$$r(\hat{w}_n(k)) - r(w^*) = (\bar{V}_n - k w^*)^2/(\bar{U}_n + 1 + k)^2 =: h(\bar{U}_n, \bar{V}_n, k), \tag{8}$$

whence $J_n(k) = \mathbb{E}\,h(\bar{U}_n, \bar{V}_n, k)$, where we have introduced the function h (used heavily in what follows) as well as the arithmetic means $\bar{U}_n$ and $\bar{V}_n$

$$\bar{U}_n := \frac{1}{n}\sum_1^n U_i, \quad \bar{V}_n := \frac{1}{n}\sum_1^n V_i, \quad \text{with} \tag{9}$$

$$U_i := X_i^2 - 1, \quad V_i := X_i Y_i - w^* X_i^2 \tag{10}$$

For convenience, also define $U := X^2 - 1$ and $V := XY - w^* X^2$. Notice that $U; U_1, U_2, \ldots$ are zero mean i.i.d. random variables, and that $V; V_1, V_2, \ldots$ satisfy the same conditions. Hence $\bar{U}_n$ and $\bar{V}_n$ converge to zero, and this leads to the idea of using the Taylor expansion of $h(u, v, k)$ about the point $(u, v) = (0, 0)$ in order to get an expansion for $J_n(k)$.

To outline the ideas, let $T_j(u,v,k)$ be the degree j Taylor polynomial of $(u,v) \mapsto h(u,v,k)$ about $(0,0)$, i.e., $T_j(u,v,k)$ is a polynomial in u and v whose coefficients are functions of k and whose degree with respect to u and v is j. Then $\mathbb{E}\,T_j(\bar{U}_n,\bar{V}_n,k)$ depends on n and moments of U and V. By deriving an upper bound for the quantity $\mathbb{E}\,|h(\bar{U}_n,\bar{V}_n,k) - T_j(\bar{U}_n,\bar{V}_n,k)|$ we get an upper bound for the error committed in approximating $J_n(k)$ by $\mathbb{E}\,T_j(\bar{U}_n,\bar{V}_n,k)$. It turns out that for odd degrees j the error is of the same order of magnitude in n as for degree $j-1$. Therefore we only consider even degrees j. It also turns out that the error bounds are uniform in $k \geq 0$ whenever $j \geq 2$. To proceed, we need to introduce assumptions.

Assumption 1 $\mathbb{E}\,|X|^r < \infty$ *and* $\mathbb{E}\,|Y|^s < \infty$ *for high enough* r *and* s.

Assumption 2 *Either (a) for some constant* $\beta > 0$ *almost surely* $|X| \geq \beta$ *or (b)* X *has a density which is bounded in some neighborhood of zero.*

Assumption 1 guarantees the existence of high enough moments; the values $r = 20$ and $s = 8$ are sufficient for the following proofs. E.g., if the pair (X,Y) has a normal distribution or a distribution with compact support, then moments of all orders exist and hence in this case assumption 1 would be satisfied. Without some condition such as assumption 2, $J_n(0)$ might fail to be meaningful or finite. The following technical result is stated without proof.

Proposition 2 *Let* $p > 0$ *and let* $0 < \mathbb{E}\,X^2 < \infty$. *If assumption 2 holds, then*

$$\mathbb{E}\left\{\left[\frac{1}{n}(X_1^2 + \cdots + X_n^2)\right]^{-p}\right\} \to \left[\mathbb{E}\,(X^2)\right]^{-p}, \quad \text{as} \quad n \to \infty,$$

where the expectation on the left is finite (a) for $n \geq 1$ *(b) for* $n > 2p$ *provided that assumption 2 (a), respectively 2 (b) holds.*

Proposition 3 *Let assumptions 1 and 2 hold. Then there exist constants* n_0 *and* M *such that*

$$\begin{aligned} J_n(k) &= \mathbb{E}\,T_2(\bar{U}_n,\bar{V}_n,k) + R(n,k) \quad \textit{where} \\ \mathbb{E}\,T_2(\bar{U}_n,\bar{V}_n,k) &= \frac{(w^*)^2k^2}{(1+k)^2} + n^{-1}\left[\frac{\mathbb{E}\,V^2}{(1+k)^2} + 3\frac{(w^*)^2k^2\mathbb{E}\,U^2}{(1+k)^4} + 4\frac{w^*k\mathbb{E}\,UV}{(1+k)^3}\right] \\ |R(n,k)| &\leq Mn^{-3/2}(k+1)^{-1}, \quad \forall n \geq n_0, k \geq 0. \end{aligned}$$

PROOF SKETCH The formula for $\mathbb{E}\,T_2(\bar{U}_n,\bar{V}_n,k)$ follows easily by integrating the degree two Taylor polynomial term by term. To get the upper bound for $R(n,k)$, consider the residual

$$h(u,v,k) - T_2(u,v,k) = \frac{-2(k+1)^3uv^2 + -4(w^*)^2k^2(k+1)u^3 + \cdots}{(u+1+k)^2(k+1)^4},$$

where we have omitted four similar terms. Using the bound

$$(\bar{U}_n + 1 + k)^2 = \left(\frac{1}{n}\sum_1^n X_i^2 + k\right)^2 \geq \left(\frac{1}{n}\sum_1^n X_i^2\right)^2, \quad \forall k \geq 0,$$

the L_1 triangle inequality, and the Cauchy-Schwartz inequality, we get

$$|R(n,k)| = |\mathbb{E}\,[h(\bar{U}_n,\bar{V}_n,k) - T_2(\bar{U}_n,\bar{V}_n,k)]|$$
$$\leq (k+1)^{-4}\left\{\mathbb{E}\left[(\frac{1}{n}\sum_1^n X_i^2)^{-4}\right]\right\}^{1/2}$$
$$\left\{2(k+1)^3[\mathbb{E}\,(|\bar{U}_n|^2|\bar{V}_n|^4)]^{1/2} + 4(w^*)^2k^2(k+1)[\mathbb{E}\,|\bar{U}_n|^6]^{1/2}\ldots\right\}$$

By proposition 2, here $\mathbb{E}\,[(\frac{1}{n}\sum_1^n X_i^2)^{-4}] = O(1)$. Next we use the following fact, cf. (Serfling, 1980, Lemma B, p. 68).

Fact 1 *Let $\{Z_i\}$ be i.i.d. with $\mathbb{E}\,[Z_1] = 0$ and with $\mathbb{E}\,|Z_1|^\nu < \infty$ for some $\nu \geq 2$. Then*

$$\mathbb{E}\left|\frac{1}{n}\sum_1^n Z_i\right|^\nu = O(n^{-\nu/2})$$

Applying the Cauchy-Schwartz inequality and this fact, we get, e.g., that

$$[\mathbb{E}\,(|\bar{U}_n|^2|\bar{V}_n|^4)]^{1/2} \leq [(\mathbb{E}\,|\bar{U}_n|^4)^{1/2}(\mathbb{E}\,|\bar{V}_n|^8)^{1/2}]^{1/2} = O(n^{-3/2}).$$

Going through all the terms carefully, we see that the bound holds. □

Proposition 4 *Let assumptions 1 and 2 hold, assume that $w^* \neq 0$, and set*

$$\alpha_1 := (\mathbb{E}\,V^2 - 2w^*\mathbb{E}\,[UV])/(w^*)^2.$$

If $\alpha_1 > 0$, then there exists a constant n_1 such that for all $n \geq n_1$ the function $k \mapsto \mathbb{E}\,T_2(\bar{U}_n,\bar{V}_n,k)$ has a unique minimum on $[0,\infty)$ at the point k_n^* admitting the expansion

$$k_n^* = \alpha_1 n^{-1} + O(n^{-2}); \quad \textit{further,}$$
$$J_n(0) - \inf\{J_n(k) : k \geq 0\} = J_n(0) - J_n(\alpha_1 n^{-1}) = \alpha_1^2(w^*)^2n^{-2} + O(n^{-5/2}).$$

If $\alpha \leq 0$, then

$$\inf\{J_n(k) : k \geq 0\} = J_n(0) + O(n^{-5/2}).$$

PROOF SKETCH The proof is based on perturbation expansion considering $1/n$ a small parameter. By the previous proposition, $S_n(k) := \mathbb{E}\,T_2(\bar{U}_n,\bar{V}_n,k)$ is the sum of $(w^*)^2k^2/(1+k)^2$ and a term whose supremum over $k \geq k_0 > -1$ goes to zero as $n \to \infty$. Here the first term has a unique minimum on $(-1,\infty)$ at $k = 0$. Differentiating S_n we get

$$S_n'(k) = [2(w^*)^2k(k+1)^2 + n^{-1}p_2(k)]/(k+1)^5,$$

where $p_2(k)$ is a second degree polynomial in k. The numerator polynomial has three roots, one of which converges to zero as $n \to \infty$. A regular perturbation expansion for this root, $k_n^* = \alpha_1 n^{-1} + \alpha_2 n^{-2} + \ldots$, yields the stated formula for α_1. This point is a minimum for all sufficiently large n; further, it is greater than zero for all sufficiently large n if and only if $\alpha_1 > 0$.

The estimate for $J_n(0) - \inf\{J_n(k) : k \geq 0\}$ in the case $\alpha_1 > 0$ follows by noticing that

$$J_n(0) - J_n(k) = \mathbb{E}\,[h(\bar{U}_n,\bar{V}_n,0) - h(\bar{U}_n,\bar{V}_n,k)],$$

where we now use a third degree Taylor expansion about $(u,v,k) = (0,0,0)$

$$h(u,v,0) - h(u,v,k) =$$
$$2w^*kv - (w^*)^2k^2 - 4w^*kuv + 2(w^*)^2k^2u + 2kv^2 - 4w^*k^2v + 2(w^*)^2k^3 + r(u,v,k).$$

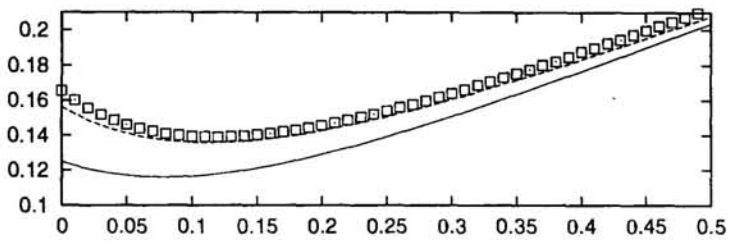


Figure 1: Illustration of the asymptotic approximations in the situation of equation (11). Horizontal axis k; vertical axis $J_n(k)$ and its asymptotic approximations. Legend: markers $J_n(k)$; solid line $\mathbb{E}\,T_2(\bar{U}_n, \bar{V}_n, k)$; dashed line $\mathbb{E}\,T_4(\bar{U}_n, \bar{V}_n, k)$.

Using the techniques of the previous proposition, it can be shown that $\mathbb{E}\,|r(\bar{U}_n, \bar{V}_n, k_n^*)| = O(n^{-5/2})$. Integrating the Taylor polynomial and using this estimate gives

$$J_n(0) - J_n(\alpha_1/n) = \alpha_1^2 (w^*)^2 n^{-2} + O(n^{-5/2}).$$

Finally, by the mean value theorem,

$$\begin{aligned} J_n(0) - \inf\{J_n(k) : k \geq 0\} &= J_n(0) - J_n(\alpha_1/n) + \frac{d}{dk}[J_n(0) - J_n(k)]_{|k=\theta}(k_n^* - \alpha_1/n) \\ &= J_n(0) - J_n(\alpha_1/n) + O(n^{-1})O(n^{-2}) \end{aligned}$$

where θ lies between k_n^* and α_1/n, and where we have used the fact that the indicated derivative evaluated at θ is of order $O(n^{-1})$, as can be shown with moderate effort. □

Remark In the preceding we assumed that $A = \mathbb{E}X^2$ equals 1. If this is not the case, then the formula for α_1 has to be divided by A; again, if $\alpha_1 > 0$, then $k_n^* = \alpha_1 n^{-1} + O(n^{-2})$.

If the model is correctly specified in the sense that $Y = w^* X + \epsilon$, where ϵ is independent of X and $\mathbb{E}\epsilon = 0$, then $V = X\epsilon$ and $\mathbb{E}[UV] = 0$. Hence we have $\alpha_1 = \mathbb{E}[\epsilon^2]/(w^*)^2$, and this is strictly positive expect in the degenerate case where $\epsilon = 0$ with probability one. This means that here regularization helps provided the regularization parameter is chosen around the value α_1/n and n is large enough. See Figure 1 for an illustration in the case

$$X \sim N(0,1), \quad Y = w^* X + \epsilon, \quad \epsilon \sim N(0,1), \quad w^* = 1, \tag{11}$$

where ϵ and X are independent. $J_n(k)$ is estimated on the basis of 1000 repetitions of the task for $n = 8$. In addition to $\mathbb{E}\,T_2(\bar{U}_n, \bar{V}_n, k)$ the function $\mathbb{E}\,T_4(\bar{U}_n, \bar{V}_n, k)$ is also plotted. The latter can be shown to give $J_n(k)$ correctly up to order $O(n^{-5/2}(k+1)^{-3})$. Notice that although $\mathbb{E}\,T_2(\bar{U}_n, \bar{V}_n, k)$ does not give that good an approximation for $J_n(k)$, its minimizer is near the minimizer of $J_n(k)$, and both of these minimizers lie near the point $\alpha_1/n = 0.125$ as predicted by the theory. In the situation (11) it can actually be shown by lengthy calculations that the minimizer of $J_n(k)$ is exactly α_1/n for each sample size $n \geq 1$.

It is possible to construct cases where $\alpha_1 < 0$. For instance, take

$$\begin{aligned} X &\sim \text{Uniform}(a,b), \quad a = \frac{1}{2}, b = \frac{1}{4}(3\sqrt{5} - 1) \\ Y &= c/X + d + Z, \quad c = -5, d = 8, \end{aligned}$$

and $Z \sim N(0, \sigma^2)$ with Z and X independent and $0 \leq \sigma < 1.1$. In such a case regularization using a positive regularization parameter only makes matters worse; using a properly chosen *negative* regularization parameter would, however, help in this particular case. This would, however, amount to rewarding rapidly changing functions. In the case (11) regularization using a negative value for the regularization parameter would be catastrophic.

4 DISCUSSION

We have obtained asymptotic approximations for the optimal regularization parameter in (1) and the amount of improvement (3) in the simple case of one-dimensional linear regression when the regularization parameter is chosen independently of the training sample. It turned out that the optimal regularization parameter is, to leading order, given by $\alpha_1 n^{-1}$ and the resulting improvement is of order $O(n^{-2})$. We have also seen that if $\alpha_1 < 0$ then regularization only makes matters worse.

Also (Larsen and Hansen, 1994) have obtained asymptotic results for the optimal regularization parameter in (1). They consider the case of a nonlinear network; however, they assume that the neural network model is correctly specified.

The generalization of the present results to the nonlinear, misspecified case might be possible using, e.g., techniques from (Bhattacharya and Ghosh, 1978). Generalization to the case where the regularization parameter is chosen on the basis of the sample (say, by cross validation) would be desirable.

Acknowledgements

This paper was prepared while the author was visiting the Department for Statistics and Probability Theory at the Vienna University of Technology with financial support from the Academy of Finland. I thank F. Leisch for useful discussions.

References

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *The Annals of Statistics*, 6(2):434–451.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer series in statistics. Springer-Verlag.

Hoerl, A. E. and Kennard, R. W. (1988). Ridge regression. In Kotz, S., Johnson, N. L., and Read, C. B., editors, *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc.

Larsen, J. and Hansen, L. K. (1994). Generalization performance of regularized neural network models. In Vlontos, J., Whang, J.-N., and Wilson, E., editors, *Proc. of the 4th IEEE Workshop on Neural Networks for Signal Processing*, pages 42–51. IEEE Press.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc.

Titterington, D. M. (1985). Common structure of smoothing techniques in statistics. *International Statistical Review*, 53:141–170.

A General Purpose Image Processing Chip: Orientation Detection

Ralph Etienne-Cummings and Donghui Cai
Department of Electrical Engineering
Southern Illinois University
Carbondale, IL 62901-6603

Abstract

A 80 x 78 pixel general purpose vision chip for spatial focal plane processing is presented. The size and configuration of the processing receptive field are programmable. The chip's architecture allows the photoreceptor cells to be small and densely packed by performing all computation on the read-out, away from the array. In addition to the raw intensity image, the chip outputs four processed images in parallel. Also presented is an application of the chip to line segment orientation detection, as found in the retinal receptive fields of toads.

1 INTRODUCTION

The front-end of the biological vision system is the retina, which is a layered structure responsible for image acquisition and pre-processing. The early processing is used to extract spatiotemporal information which helps perception and survival. This is accomplished with cells having feature detecting receptive fields, such as the edge detecting center-surround spatial receptive fields of the primate and cat bipolar cells [Spillmann, 1990]. In toads, the receptive fields of the retinal cells are even more specialized for survival by detecting "prey" and "predator" (from size and orientation filters) at this very early stage [Spillmann, 1990].

The receptive of the retinal cells performs a convolution with the incident image in parallel and continuous time. This has inspired many engineers to develop retinomorphic vision systems which also imitate these parallel processing capabilities [Mead, 1989; Camp, 1994]. While this approach is ideal for fast early processing, it is not space efficient. That is, in realizing the receptive field within each pixel, considerable die area is required to implement the convolution kernel. In addition, should programmability be required, the complexity of each pixel increases drastically. The space constraints are eliminated if the processing is performed serially during read-out. The benefits of this approach are 1) each pixel can be as small as possible to allow high resolution imaging, 2) a single processor unit is used for the entire retina thus reducing mis-match problems, 3) programmability can be obtained with no impact on the density of imaging array, and

4) compact general purpose focal plane visual processing is realizable. The space constrains are then transformed into temporal restrictions since the scanning clock speed and response time of the processing circuits must scale with the size of the array. Dividing the array into sub-arrays which are scanned in parallel can help this problem. Clearly this approach departs from the architecture of its biological counterpart, however, this method capitalizes on the main advantage of silicon which is its speed. This is an example of mixed signal neuromorphic engineering, where biological ideas are mapped onto silicon not using direct imitation (which has been the preferred approach in the past) but rather by realizing their *essence* with the best silicon architecture and computational circuits.

This paper presents a general purpose vision chip for spatial focal plane processing. Its architecture allows the photoreceptor cells to be small and densely packed by performing all computation on the read-out, away from the array. Performing computation during read-out is ideal for silicon implementation since no additional temporal over-head is required, provided that the processing circuits are fast enough. The chip uses a single convolution kernel, per parallel sub-array, and the scanning bit pattern to realize various receptive fields. This is different from other focal plane image processors which and usually restricted to hardwired convolution kernels, such as oriented 2D Gabor filters [Camp, 1994]. In addition to the raw intensity image, the chip outputs four processed versions per sub-array. Also presented is an application of the chip to line segment orientation detection, as found in the retinal receptive fields of toads [Spillmann, 1990].

2 THE GENERAL PURPOSE IMAGE PROCESSING CHIP

2.1 System Overview

This chip has an 80 row by 78 column photocell array partitioned into four independent sub-arrays, which are scanned and output in parallel, (see figure 1). Each block is 40 row by 39 column, and has its own convolution kernel and output circuit. The scanning circuit includes three parts: virtual ground, control signal generator (CSG), and scanning output transformer. Each block has its own virtual ground and scanning output transformer in both x direction (horizontal) and y direction (vertical). The control signal generator is shared among blocks.

2.2 Hardware Implementation

The photocell is composed of phototransistor, photo current amplifier, and output control. The phototransistor performance light transduction, while the amplifier magnifies the photocurrent by three orders of magnitude. The output control provides multiple copies of the amplified photocurrent which is subsequently used for focal plane image processing.

The phototransistor is a parasitic PNP transistor in an Nwell CMOS process. The current amplifier uses a pair of diode connected pmosfets to obtain a logarithmic relationship between light intensity and output current. This circuit also amplifies the photocurrent from nanoamperes to microamperes. The photocell sends three copies of the output currents into three independent buses. The connections from the photocell to the buses are controlled by pass transistors, as shown in Fig. 2. The three current outputs allow the image to be processed using multiple receptive field organization (convolution kernels), while the raw image is also output. The row (column) buses provides currents for extracting horizontally (vertically) oriented image features, while the original bus provides the logarithmically compressed intensity image.

The scanning circuit addresses the photocell array by selecting groups of cells at one time. Since the output of the cells are currents, virtual ground circuits are used on each bus to mask the > 1pF capacitance of the buses. The CSG, implemented with shift registers

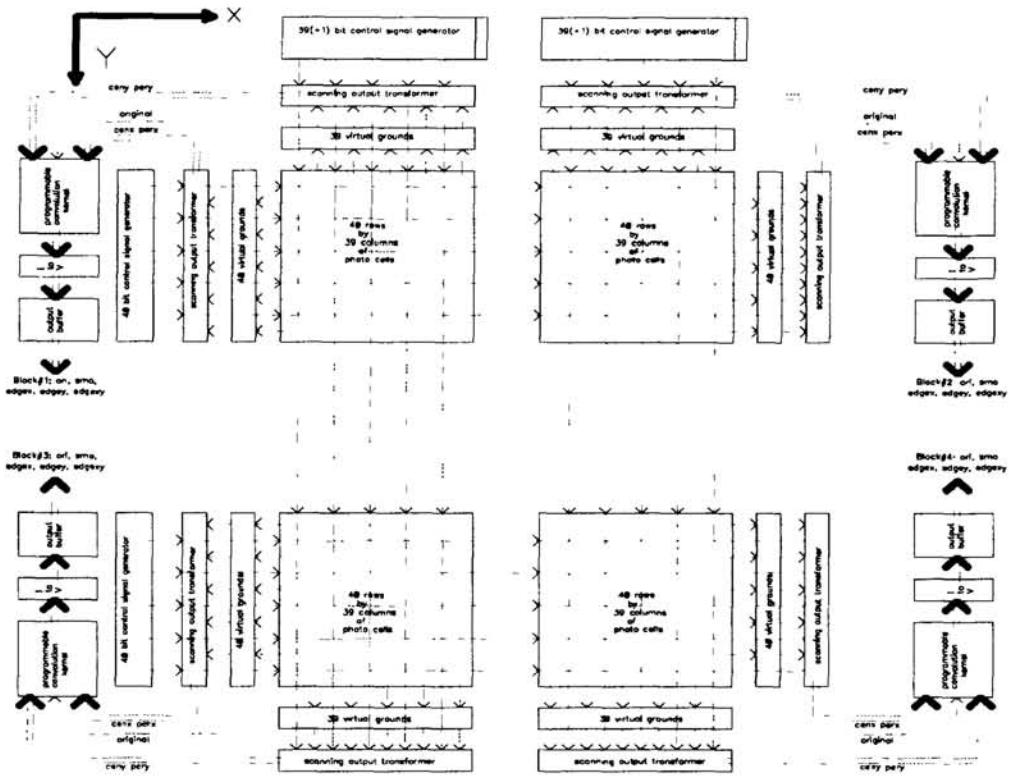


Figure 1: Block diagram of the chip.

produces signals which select photocells and control the scanning output transformer. The scanning output transformer converts currents from all row buses into I_{perx} and I_{cenx}, and converts currents from all row buses into I_{pery} and I_{ceny}. This transformation is required to implement the various convolution kernels discussed later.

The output transformer circuits are controlled by a central CSG and a peripheral CSG. These two generators have identical structures but different initial values. It consists of an n-bit shift register in x direction (horizontally) and an m-bit shift register in y direction (vertically). A feedback circuit is used to restore the scanning pattern into the x shift register after each row scan is completed. This is repeated until all the row in each block are scanned.

The control signals from the peripheral and central CSGs select all the cells covered by a 2D convolution mask (receptive field). The selected cells send I_{xy} to the original bus, I_{xp} to the row bus, and I_{yp} to the column bus. The function of the scanning output transformer is to identify which rows (columns) are considered as the center (I_{cenx} or I_{ceny}) or periphery (I_{perx} or I_{pery}) of the convolution kernel, respectively. Figure 3 shows how a 3x3 convolution kernel can be constructed.

Figure 4 shows how the output transformer works for a 3x3 mask. Only row bus transformation is shown in this example, but the same mechanism applies to the column bus as well. The photocell array is m row by n column, and the size is 3x3. The XC (x center) and YC (y center) come from the central CSG; while XP (x peripheral) and YP (y peripheral) come from the peripheral CSG. After loading the CSG, the initial values of XP and YP are both 00011...1. The initial values of XC and YC are both 10111...1. This identifies the central cell as location (2, 2). The currents from the central row (column) are summed to form I_{cenx} and I_{ceny}, while all the peripheral cells are summed to form I_{perx} and I_{pery}. This is achieved by activating the switches labeled XC, YC, XP and YP in figure 2. XP_i (YP_i) {i=1, 2, ..., n} controls whether the output current of one cell

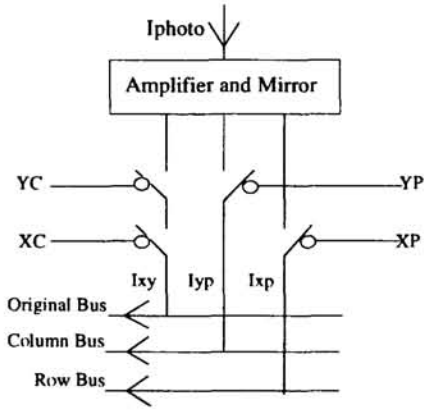


Figure 2: Connections between a photo-cell and the current buses.

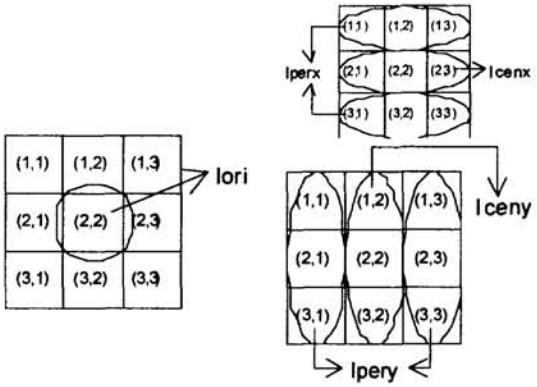


Figure 3: Constructing a 3x3 receptive field.

goes to the row (column) bus. Since XP_i (YP_i) is connected to the gate of a pmos switch, a 0 in XP_i (YP_i) it turns on. YC_i (XC_i) {i=1, 2, ..., n} controls whether a row (column) bus connects to I_{cenx} bus in the same way. On the other hand, the connection from a row (column) bus to I_{perx} bus is controlled by an nmos and a pmos switch. The connection is made if and only if YC_i (XC_i), an nmos switch, is 1 and YP_i (XP_i), a pmos switches, is 0. The intensity image is obtained directly when XC_i and YC_i are both 0. Hence, $I_{ori} = I(2,2)$, $I_{cenx} = I_{row2} = I(2,1) + I(2,2) + I(2,3)$ and $I_{perx} = I_{row1} + I_{row3} = I(1,1) + I(1,2) + I(1,3) + I(3,1) + I(3,2) + I(3,3)$.

The convolution kernel can be programmed to perform many image processing tasks by loading the scanning circuit with the appropriate bit pattern. This is illustrated by configuring the chip to perform image smoothing and edge extraction (x edge, y edge, and 2D edge), which are all computed simultaneously on read-out. It receives five inputs (I_{ori}, I_{cenx}, I_{perx}, I_{ceny}, I_{pery}) from the scanning circuit and produces five outputs (I_{ori}, I_{edgex}, I_{edgey}, I_{smooth}, I_{edge2d}). The kernel (receptive field) size is programmable from 3x3, 5x5, 7x7, 9x9 and 11x11. Fig. 5 shows the 3x3 masks for this processing. Repeating the above steps for 5x5, 7x7, 9x9, and 11x11 masks, we can get similar results.

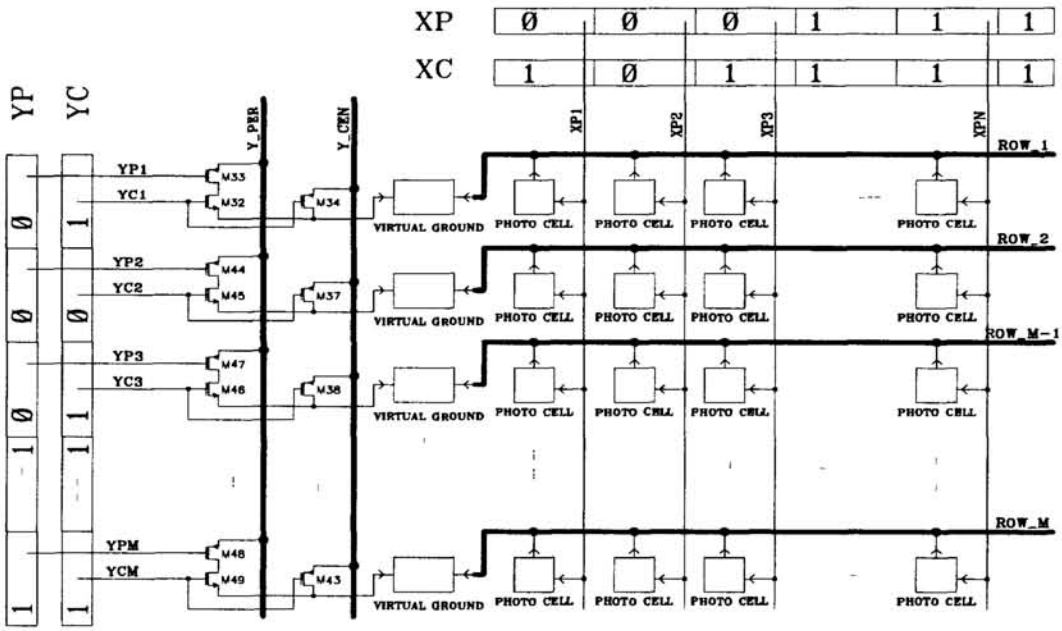


Figure 4: Scanning output transformer for an m row by n column photo cell array.

$I_{smooth}=I_{cenx}+I_{perx}$ $I_{edgex}=2*I_{cenx}-I_{perx}$ $I_{edgey}=2*I_{ceny}-I_{pery}$ $I_{edge2D}=6*I_{ori}-I_{cenx}-I_{ceny}$

1	1	1
1	1	1
1	1	1

(a) smooth

-1	-1	-1
2	2	2
-1	-1	-1

(b) edge_x

-1	2	-1
-1	2	-1
-1	2	-1

(c) edge_y

0	-1	0
-1	4	-1
0	-1	0

(d) edge_2D

Figure 5: 3x3 convolution masks for various image processing.

In general, the convolution results under different mask sizes can be expressed as follows:

$I_{smooth}=I_{cenx}+I_{perx}$ $I_{edgex}=K_{1d}*I_{cenx}-I_{perx}$ $I_{edgey}=K_{1d}*I_{ceny}-I_{pery}$ $I_{edge2D}=K_{2d}*I_{ori}-I_{cenx}-I_{ceny}$

Where K_{1d} and K_{2d} are the programmable coefficients (from 2-6 and 6-14, respectively) for 1D edge extraction and 2D edge extraction, respectively. By varying the locations of the 0's in the scanning circuits, different types of receptive fields (convolution kernels) can be realized.

2.3 Results

The chip contains 65K transistors in a footprint of 4.6 mm x 4.7 mm. There are 80 x 78 photocells in the chip, each of which is 45.6 μm x 45 μm and a fill factor of 15%. The convolution kernel occupies 690.6 μm x 102.6 μm. The power consumption of the chip for a 3x3 (11x11) receptive field, indoor light, and 5V power supply is < 2 mW (8 mW).

To capitalize on the programmability of this chip, an A/D card in a Pentium 133MHz PC is used to load the scanning circuit and to collect data. The card, which has a maximum analog throughput of 100 KHz limits the frame rate of the chip to 12 frames per second. At this rate, five processed versions of the image is collected and displayed. The scanning and processing circuits can operate at 10 MHz (6250 fps), however, the phototransistors have much slower dynamics. Temporal smoothing (smear) can be observed on the scope when the frame rate exceeds 100 fps.

The chip displays a logarithmic relationship between light intensity and output current (unprocessed imaged) from 0.1 lux (100 nA) to 6000 lux (10 μA). The fixed pattern noise, defined as standard-deviation/mean, decreases abruptly from 25% in the dark to 2% at room light (800 lux). This behavior is expected since the variation of individual pixel current is large compared to the mean output when the mean is small. The logarithmic response of the photocell results in high sensitivity at low light, thus increasing the mean value sharply. Little variation is observed between chips.

The contrast sensitivity of the edge detection masks is also measured for the 3x3 and 5x5 receptive fields. Here contrast is defined as $(I_{max} - I_{min})/(I_{max} + I_{min})$ and sensitivity is given as a percentage of the maximum output. The measurements are performed for normal room and bright lighting conditions. Since the two conditions corresponded to the saturated part of the logarithmic transfer function of the photocells, then a linear relationship between output response and contrast is expected. Figure 6 shows contrast sensitivity plot. Figure 7 shows examples of chip's outputs. The top two images are the raw and smoothed (5x5) images. The bottom two are the 1D edge_x (left) and 2D edge (right) images. The pixels with positive values have been thresholded to white. The vertical black line in the image is not visible in the edge_x image, but can be clearly seen in the edge_2D image.

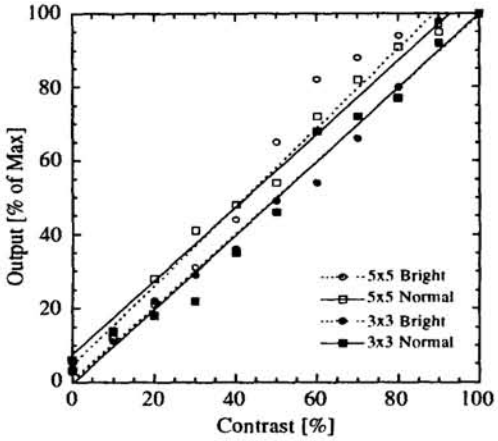


Figure 6: Contrast sensitivity function of the x edge detection mask.

Figure 7: (Clockwise) Raw image, 5x5 smoothed image, edge_2D and edge_x.

3 APPLICATION: ORIENTATION DETECTION

3.1 Algorithm Overview

This vision chip can be elegantly used to measure the orientation of line segments which fall across the receptive field of each pixel. The output of the 1D Laplacian operators, edge_x and edge_y, shown in figure 5, can be used to determine the orientation of edge segments. Consider a continuous line through the origin, represented by a delta function in 2D space by $\delta(y-x\tan\theta)$. If the origin is the center of the receptive field, the response of the edge_x kernel can be computed by evaluating the convolution equation (1), where $W(x) = u(x+m)-u(x-m)$ is the x window over which smoothing is performed, $2m+1$ is the width of the window and $2n+1$ is the number of coefficients realizing the discrete Laplacian operator. In our case, $n = m$. Evaluating this equation and substituting the origin for the pixel location yields equation (2), which indicates that the output of the 1D edge_x (edge_y) detectors have a discretized linear relationship to orientation from 0° to 45° (45° to 90°). At 0°, the second term in equation (2) is zero. As θ increase, more terms are subtracted until all terms are subtracted at 45°. Above 45° (below 45°), the edge_x (edge_y) detectors output zero since equal numbers of positive and negative coefficients are summed. Provided that contrast can be normalized, the output of the detectors can be used to extract the orientation of the line. Clearly these responses are even about the x- and y-axis, respectively. Hence, a second pair of edge detectors, oriented at 45°, is required to uniquely extract the angle of the line segment.

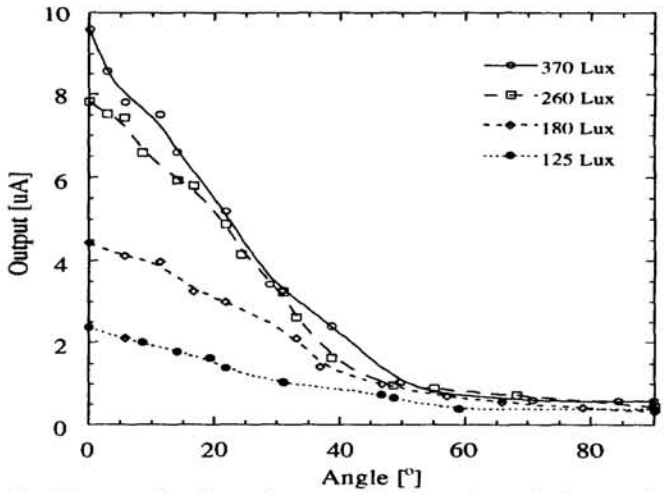


Figure 8: Measured orientation transfer function of edge_x detectors.

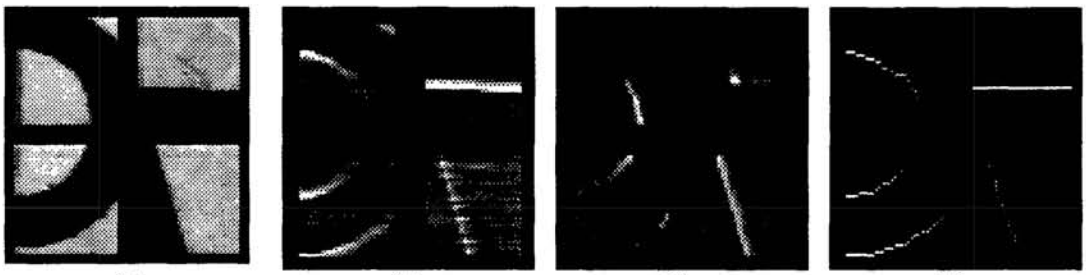
$$O_{edge\_x}(x,y) = [\,2nW(x \pm m)\delta(y) - \sum_{i=1}^{n} W(x \pm m)\delta(y \pm i)\,] * \delta(y - x\tan\theta) \qquad (1)$$

$$O_{edge\_x}(0,0) = 2n - [\sum_{i=1}^{n}(W(\frac{i}{\tan\theta}) + W(\frac{-i}{\tan\theta})] \qquad (2)$$

3.2 Results

Figure 8 shows the measured output of the edge_x detectors for various lighting conditions as a line is rotated. The average positive outputs are plotted. As expected, the output is maximum for bright ambients when the line is horizontal. As the line is rotated, the output current decreases linearly and levels off at approximately 45°. On the other hand, the edge_y (not shown) begins its linear increase at 45° and maximizes at 90°. After normalizing for brightness, the four curves are very similar (not shown).

To further demonstrate orientation detection with this chip, a character consisting of a circle and some straight lines is presented. The intensity image of the character is shown in figure 9(a). Figures 9(b) and 9(c) show the outputs of the edge_x and edge_y detectors, respectively. Since a 7x7 receptive field is used in this experiment, some outer pixels of each block are lost. The orientation selectivity of the 1D edge detectors are clearly visible in the figures, where edge_x highlights horizontal edges and edge_y vertical edges. Figure 9(d) shows the reported angles. A program is written which takes the two 1D edge images, finds the location of the edges from the edge_2D image, the intensity at the edges (positive lobe) and then computes the angle of the edge segment. In figure 9(d), the black background is chosen for locations where no edges are detected, white is used for 0° and gray for 90°.



(a) (b) (c) (d)

Figure 9: Orientation detection using 1D Laplacian Operators.

4 CONCLUSION

A 80x78 pixel general purpose vision chip for spatial focal plane processing has been presented. The size and configuration of the processing receptive field are programmable. In addition to the raw intensity image, the chip outputs four processed images in parallel. The chip has been successfully used for compact line segment orientation detection, which can be used in character recognition. The programmability and relatively low power consumption makes it ideal for many visual processing tasks.

References

Camp W. and J. Van der Spiegel, "A Silicon VLSI Optical Sensor for Pattern Recognition, " *Sensors and Actuators A*, Vol. 43, No. 1-3, pp. 188-195, 1994.

Mead C. and M. Ismail (Eds.), *Analog VLSI Implementation of Neural Networks*, Kluwer Academic Press, Newell, MA, 1989.

Spillmann L. and J. Werner (Eds.), *Visual Perception: The Neurophysiological Foundations*, Academic Press, San Diego, CA, 1990.