Ellipsoid Projection onto Halfspaces

Ezra Winston

April 2019

Contents

1 Bounding Ellipsoid Projection onto Halfspace

We are interested the projection of an ellipsoid $E = \mathcal{E}(q, Q)$ onto a halfspace $\mathcal{H}(c, \gamma)$ where

$$\mathcal{E}(q, Q) = \{x : (x - q)^\top Q^{-1}(x - q) \leq 1\} \quad \text{and} \quad \mathcal{H}(c, \gamma) = \{x : x^\top c \leq \gamma\}.$$

To approximate this, we would like to find the min. vol. ellipsoid covering the projection, $E^* = MVE(Proj_H(E))$.

There are five cases (see Figure 1). The first two cases are easy:

Case 1 E is entirely inside H. Then

$$\begin{aligned} E^* &= MVE(Proj_H(E)) \\ &= MVE(E) \\ &= E. \end{aligned}$$

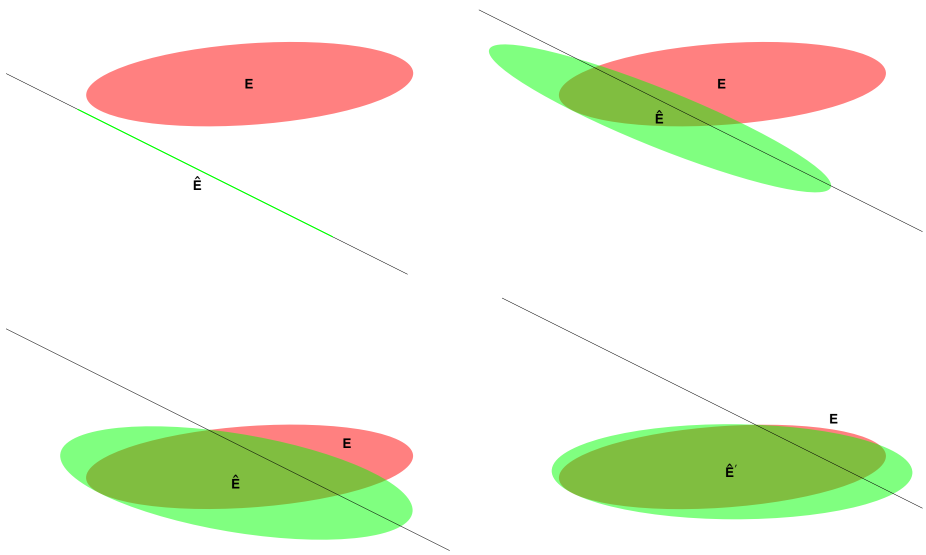

Figure 1: Upper left: Case 2; Upper right: Case 3; Lower left: Case 4; Lower right: Case 5.

Case 2 E is entirely outside H. Then

$$\begin{aligned} E^* &= MVE(Proj_H(E)) \\ &= MVE(Proj_{HP}(E)) \\ &= Proj_{HP}(E), \end{aligned}$$

where HP is the hyperplane $HP = \{x : x^\top c = \gamma\}$. The projection of an ellipsoid onto a hyperplane is another ellipsoid which can be computed following Fact 2.

Cases 3-5 deal with the scenario when E intersects both H and H^c. We can borrow techniques from the ellipsoid method [1] for these cases. The cases split on the value α, the distance from q to the hyperplane HP in the metric defined by Q (i.e., $\|y\|_Q = (y^\top Q^{-1} y)^{1/2}$). We compute α as $\alpha = (c^\top q - \gamma)/\sqrt{c^\top Qc}$. Note that Case 1 is when $\alpha < -1$ and Case 2 is when $\alpha > 1$.

Note that

$$E^* = MVE(Proj_{HP}(E \cap H^c) \cup (E \cap H)).$$

In Cases 3 and 4, since $Proj_{HP}(E \cap H^c)$ is not an ellipsoid, we will compute an over-approximation to E^*. The over-approximation is

$$\hat{E} = MVE(Proj_{HP}(MVE(E \cap H^c)) \cup (E \cap H)).$$

Case 3 E intersects H and H^c and $1/n < \alpha < 1$. Then $MVE(E \cap H^c) = E$, i.e. there is no ellipsoid smaller than E itself which covers it's intersection with H (See Prop. **??** TODO), so

$$\begin{aligned}\hat{E} &= MVE(Proj_{HP}(MVE(E \cap H^c)) \cup (E \cap H)) \\ &= MVE(Proj_{HP}(E) \cup (E \cap H)).\end{aligned}$$

This can be computed using Theorem 1.

Case 4 E intersects H and H^c $with -1/n < \alpha < 1/n$. Then $MVE(E \cap H^c) \neq E$ as for Case 3, but $MVE(E \cap H^c)$ can be computed following the ellipsoid method (Fact 7). So $\hat{E}$ can be computed in a method similar to Case 3, given in Theorem 2.

In the final case, it is no longer easy to compute $\hat{E}$, so we comnpute a covering ellipsoid which is not the MVE.

Case 5 E intersects H and H^c and $-1 < \alpha < -1/n$. Then we compute $\hat{E}'$ which contains $Proj_{HP}(MVE(E \cap H^c)) \cup (E \cap H)$ using Theorem 3.

2 Ellipsoid Updates

The updates to ellipsoid matrix Q can all be represented in the form

$$\rho v_1 v_1^\top + \phi(I - v_2 v_2^\top)(\psi v_3 v_3^\top + \omega Q)(I - v_2 v_2^\top)$$

where constants ρ, ϕ, ψ, ω and vectors $v_1, \ldots, v_3$ are given in Tables **??** and **??**.

Notation used in the tables:

$$\begin{aligned}\delta(x) &:= (1 - x^2)n^2/(n^2 - 1) \\ \sigma(x) &:= 2(1 + nx)/((n + 1)(1 + x)) \\ \tau(x) &:= (1 + nx)/(n + 1) \\ s &:= \sqrt{c^\top Qc} \\ H &:= I - cc^\top\end{aligned}$$

Case	ρ	ϕ	ψ	ω
1	0	1	0	1
2	0	1	0	1
3	$\delta(0)(1 - \sigma(0))$	$\delta(0)$	0	1
4	$\delta(0)(1 - \sigma(0))$	$\delta(0)$	$-\delta(-\alpha)\sigma(-\alpha)/(c^\top Qc)$	$\delta(-\alpha)$
5	$1/(1 - \alpha)^2$	$1/(1 - \alpha^2)$	$-\delta(-\alpha)\sigma(-\alpha)/(c^\top Qc)$	$\delta(-\alpha)$

Case	v_1	v_2	v_3
1	0	0	0
2	0	c	0
3	$(c^\top q - s - \gamma)c - HQc/s$	c	0
4	$q - Qc/s - H(q + \tau(-\alpha)Qc/s) - \gamma c$	c	Qc
5	$q - Qc/s - H(q + \tau(-\alpha)Qc/s) - \gamma c$	c	Qc

The updates for the center q are given in Table **??**.

Case	new q
1	q
2	$Hq + \gamma c$
3	$Hq + \gamma c + \tau(0)v_1$
4	$H(q + \tau(-\alpha)Qc/s) + \gamma c + \tau(0)v_1$
5	$H(q + \tau(-\alpha)Qc/s) + \gamma c - \alpha v_1/(1 - \alpha)$

3 Method for Case 3

Algorithm 1 Case 3

Input: ellipsoid matrix Q, ellipsoid center q, halfspace normal c, halfspace bias γ, dimension n

s $\leftarrow \sqrt{c^\top Qc}$ ▷ s is required to compute α so could be passed in

$\ell \leftarrow c^\top q - s - \gamma$

$H^- \leftarrow I - cc^\top$

$z \leftarrow \ell c - H^- Qc/s$ ▷ Qc is required to compute α so could be passed in

$q' \leftarrow H^- q + \gamma c$

$\sigma \leftarrow 2/(n+1)$

$\delta \leftarrow n^2/(n^2 - 1)$

$\tau \leftarrow 1/(n+1)$

$q'' \leftarrow q' + \tau z$

$Q' \leftarrow \delta((1-\sigma)zz^\top + H^- Q H^-)$

return q'', Q'

See Algorithm 3. In Case 3, we want the MVE covering both the intersection $E \cap H$ and the projection $Proj_{HP}(E)$. Theorem 1 gives an algorithm for computing it. Here we describe the idea of the algorithm corresponding to the steps in Theorem 1. See Figure 3.

- (Steps 1-3) Compute the projection of E onto the hyperplane HP which divides H and H^c. The projection is a new $(n-1)$-dimensional ellipsoid $B = \mathcal{E}(q', Q')$.

- (Steps 4-7) As a consequence of the ellipsoid method (Fact 7), the MVE of $E \cap H$, denoted A, is tangent to E at a point r. Modify the projection B into an n-dimensional ellipsoid E' such that $E' \cap HP = B$ and E' is also tangent to E and A at r.

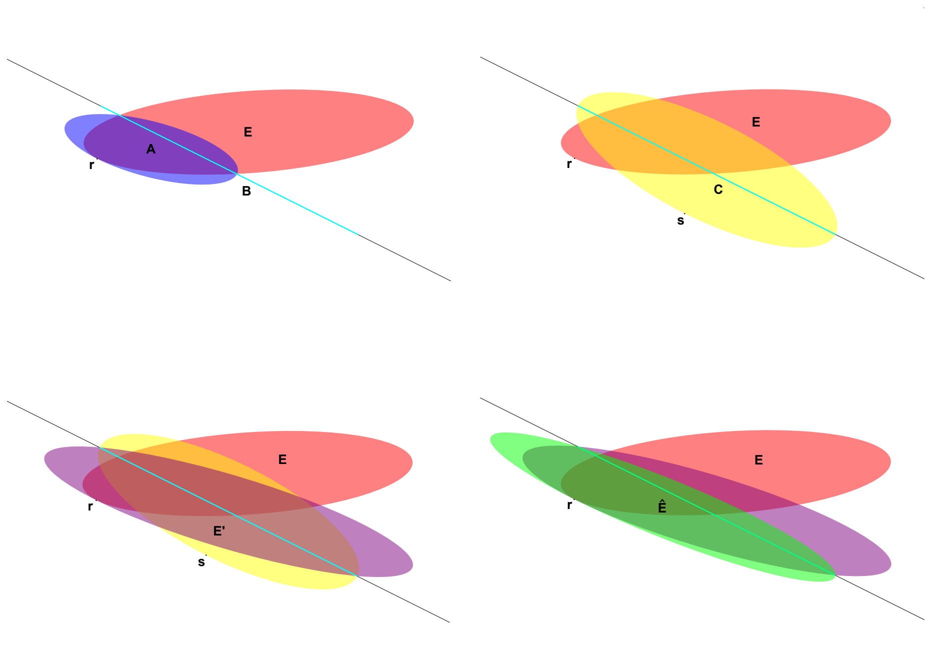

Figure 2: Steps for Case 3 method.

- (Step 7) We can show that the desired ellipsoid $\hat{E}$ is the MVE of the intersection of E' with H. This can be computed following the the ellipsoid method (Fact 7).

Note that we only care to do this computation when $1/n < \alpha < 1$, but the theorem applies to a wider range of α.

Theorem 1. *Let ellipsoid $E = \mathcal{E}(q, Q)$ and halfspace $H = \{x : c^\top x \leq \gamma\}$. Let $Proj_{HP}$ denote orthogonal projection onto $HP = \{x : c^\top x \leq \gamma\}$, where $\|c\|_2 = 1$. If $-1 < \alpha < 1$ then we can compute an ellipsoid $\hat{E}'$ which contains $(E \cap H) \cup Proj_{HP}(MVE(E \cap H^c)$ as follows:*

1. *Let P be the projection matrix defined in Fact 2.*
2. *Let $q' = PP^\top(q - \gamma c) + \gamma c$. and $Q' = PP^\top QPP^\top$.*
3. *The projection $Proj_{HP}(E)$ is $B = \mathcal{E}(q', Q')$.*
4. *Let $r = q - Qc/\sqrt{c^\top Qc}$ and $\ell = -\|r - (PP^\top(r - \gamma c) + \gamma c)\|$ and $s = q' + \ell c$.*
5. *Let S be the shear matrix mapping s to r as given in Fact 5:*
$$S = I + c((r - q') - (s - q'))^\top / ((s - q')^\top c).$$
6. *Let $Y = \ell^2 cc^\top + Q'$ and define the un-sheared ellipsoid $C = \mathcal{E}(q', Y)$.*

7. *Let $Z = S^\top Y S$ and define $E' = \mathcal{E}(q', Z)$. TODO: why S transposed?*
8. *The MVE $\hat{E}$ is the minimum volume ellipsoid for the intersection of E' and H, as computed using Fact 7.*

The proof relies on facts from Section 7 and the following propositions:

Proposition 1. *Let E be an axis-aligned ellipsoid $E = \mathcal{E}(0, diag(\lambda))$ with $\lambda_1 = 1$, and a family of parallel hyperplanes $HP_\gamma = \{x : x^\top e_1 = \gamma\}$. The intersection $E \cap HP_0$ is an also an ellipsoid. Denote it's ellipsoid matrix by Q_0. Then for $-1 < \gamma < 1$ the ellipsoid $E \cap HP_\gamma$ is similar, with matrix for $H_\gamma = (1-\gamma^2)Q_0$.*

Proof. $E \cap HP_0$ is an ellipsoid defined by

$$\sum_{i=2}^{n} \frac{x_i^2}{\lambda_i} \le 1$$

and $E \cap HP_\gamma$ is defined by

$$\sum_{i=2}^{n} \frac{x_i^2}{\lambda_i} + \gamma^2 \le 1 \qquad \Longrightarrow \qquad \sum_{i=2}^{n} \frac{x_i^2}{(1-\gamma^2)\lambda_i} \le 1.$$

So

$$Q_\gamma = \text{diag}(\frac{1}{(1-\gamma^2)\lambda_{[2:]}})^{-1} = (1-\gamma^2)\text{diag}(\lambda_{[2:]}) = (1-\gamma^2)Q_0.$$

□

Proposition 2. *Given ellipsoid $E = \mathcal{E}(q, Q)$, define*

$$p^* = \max_{x \in E} \|x\| \qquad \text{and} \qquad p^*_\delta = \max_{x \in E_\delta} \|x\|$$

where $E_\delta = \mathcal{E}(\delta q, Q)$ for some $0 \le \delta \le 1$. If $p^* \le 1$ then $p^*_\delta \le 1$.

Proof. Let a solution to the first maximization be x^* and the other be x^*_δ. First, we show that there is a solution $x^* = q + y^*$ such that $0 \le q^\top y^*$ (and the same holds for x^*_δ). To see this, note that ellipsoids are symmetric about any hyperplane through the center: If $q + y \in E$ then $q - y \in E$. Therefore, for any y with $q + y \in E$ such that $q^\top y \le 0$, then $q - y \in E$, and $0 \le q^\top(-y)$, and

$$\begin{aligned} \|q + y\|^2 &= q^\top q + 2y^\top q + y^\top y \\ &\le q^\top q - 2y^\top q + y^\top y \\ &= \|q - y\|^2. \end{aligned}$$

Next, we note that for any vector $q + y$ such that $0 \le q^\top y$, we have

$$\begin{aligned} \|\delta q + y\|^2 &= \delta^2 q^\top q + 2\delta q^\top y + y^\top y \\ &\le q^\top q + 2q^\top y + y^\top y \\ &= \|q + y\|^2. \end{aligned}$$

Since the norm of all vectors that could be the optimal solution decrease in E_δ, the norm of the optimal solution x^*_δ must decrease. □

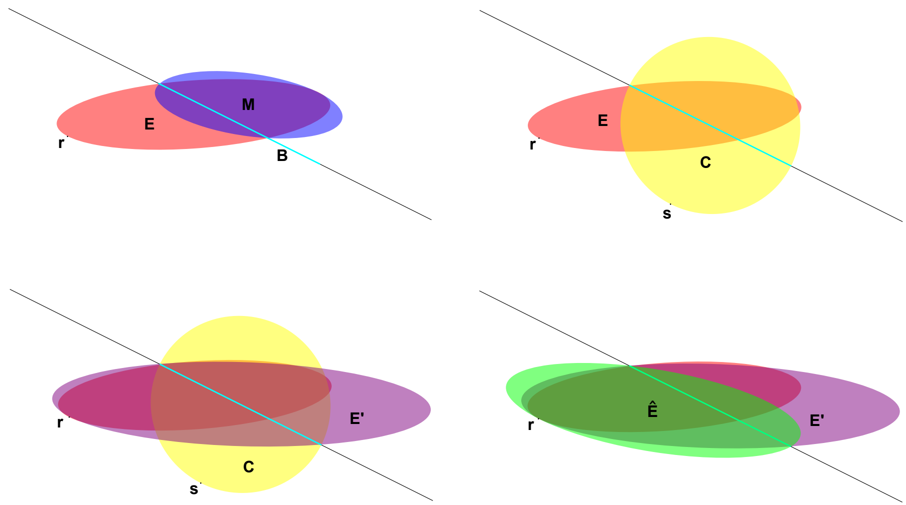

Figure 3: Steps for Case 4 method.

4 Method for Case 4

The method for Case 4 is similar to that for Case 3, but the MVE of $E \cap H^c$ is no longer E itself when $-1/n < \alpha$. Instead we compute $M = MVE(E \cap H^c)$ and then follow the steps of Theorem 1 with M in place of E. See Figure 4.

Theorem 2. *Let ellipsoid $E = \mathcal{E}(q, Q)$ and halfspace $H = \{x : c^\top x \leq \gamma\}$. Let $Proj_{HP}$ denote orthogonal projection onto $HP = \{x : c^\top x \leq \gamma\}$, where $\|c\|_2 = 1$. If $-1 < \alpha < 1/n$ then the min. volume ellipsoid $\hat{E}$ containing $(E \cap H) \cup Proj_{HP}(MVE(E \cap H^c))$ is constructed as follows: First let $M = \mathcal{E}(q_m, Q_m) = MVE(E \cap H^c)$ as computed using Fact 7. Follow the steps of Theorem 1 but replacing q with q_m and Q with Q_m.*

5 Method for Case 5

The approach taken for Case 3 and 4 is not applicable to Case 5. For Case 3, we found an ellipsoid E' passing through the projection B and the point r. We then could use the ellipsoid method update to find $\hat{E}$, the MVE covering E' and the point r.

In order to show that $\hat{E}$ contains $E \cap H$, we needed the following fact from the proof of the ellipsoid method found in [1]: For $-1/n < \alpha$, the MVE which contains $Proj_{HP}(E)$ and r also contains $MVE(E \cap H)$. This is no longer true when $\alpha < -1/n$, in which case the MVE containing $E \cap H$ is E itself (See Prop **??** TODO).

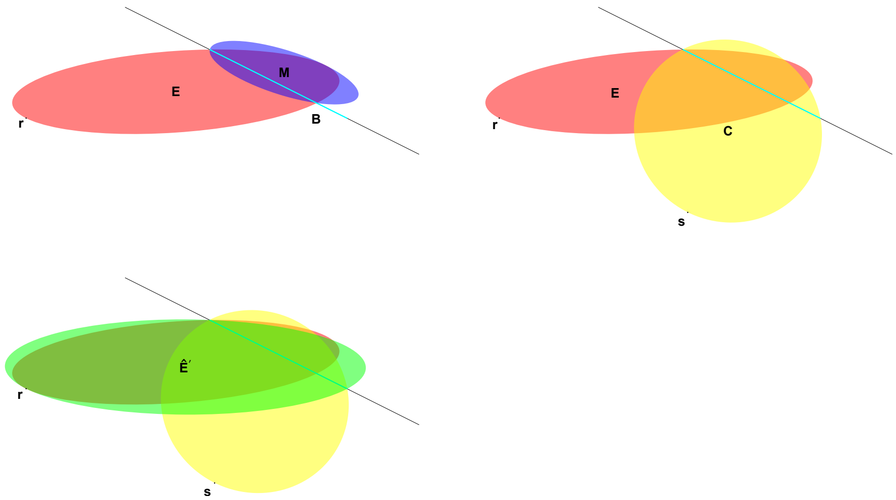

Figure 4: Steps for Case 5 method.

We can however find the ellipsoid $\hat{E}'$ that passes through through the projection $Proj_{HP}(MVE(E \cap H^c))$ and is tangent to the same hyperplane as E at r, and which has the same α as E. We can show that $E \cap H \subseteq \hat{E}'$. See Figure 5. The method is similar to that of the other cases, but instead of computing C with $\alpha = 0$, here C has the same α as E. And instead of computing the MVE for $E' \cap H$, the desired ellipsoid is E' itself, (here denoted $\hat{E}'$).

Theorem 3. *Let ellipsoid $E = \mathcal{E}(q, Q)$ and halfspace $H = \{x : c^\top x \leq \gamma\}$. Let $Proj_{HP}$ denote orthogonal projection onto $HP = \{x : c^\top x \leq \gamma\}$, where $\|c\|_2 = 1$. If $-1/n < \alpha < 1$ then the min. volume ellipsoid $\hat{E}$ containing $(E \cap H) \cup Proj_{HP}(E)$ is constructed as follows:*

1. *Let $M = M = \mathcal{E}(q_m, Q_m) = MVE(E \cap H^c)$ as computed using Fact 7.*
2. *Let P be the projection matrix defined in Fact 2.*
3. *Let $q_1' = PP^\top(q_m - \gamma c) + \gamma c$. and $Q' = PP^\top Q_m PP^\top$.*
4. *The projection $Proj_{HP}(MVE(E \cap H^c))$ is $B = \mathcal{E}(q_1', Q')$.*
5. *Let $r = q - Qc/\sqrt{c^\top Qc}$ and $\ell = -\|r - (PP^\top(r - \gamma c) + \gamma c)\|/(1 - \alpha)$.*
6. *Let $q_2' = q_1' - \alpha\ell c$ and $s = q_2' + \ell c$.*
7. *Let S be the shear matrix mapping s to r as given in Fact 5:*
$$S = I + c((r - q') - (s - q'))^\top/((s - q')^\top c).$$
8. *Let $Y = \ell^2 cc^\top + Q'/(1 - \alpha^2)$ and define the un-sheared ellipsoid $C = \mathcal{E}(q_2', Y)$.*

9. *Let $Z = S^\top Y S$ and $q_3' = S(q_2' - q_1') + q_1'$ and define $\hat{E}' = \mathcal{E}(q_3', Z)$. TODO: why S transposed?*

Proof. TODO ☐

6 Efficient Computation Using Factored Matrices

Rather than keeping track of the ellipsoid matrix Q through a network, we can keep track of A such that $Q = AA^\top$. This improves numerical stability (forcing Q to be PSD), and reduces the number of matrix multiplications required. This factorization can be maintained through the projections, and all the matrix multiplications required for halfspace projections can be computed as matrix-vector multiplications:

- The projection matrix is of the form $P = \left(I - \frac{2}{\|a\|^2} aa^\top\right)_{[:,2:]}$, so $A^\top P$ can be done with matrix-vector multiplication.
- The computations required in Step 6 of Theorem 1 can be done using the factor only: We need B such that $BB^\top = \ell cc^\top + AA^\top$, which is given by $B = [\ell c^\top; A]$.
- The sheering matrix S is of the form $I + xy^\top$ for vectors $x,\ y$, so $A^\top S$ is matrix-vector.
- Calculating MVEs following the ellipsoid method (Fact 7) involves computing $Qc/\sqrt{c^\top Qc}$ and $Q - \sigma Qc(Qc)^\top/(c^\top Qc)$. Both can be done with only matrix-vector multiplications. The later can be done in factored form as follows:
  - $Q - \sigma Qc(Qc)^\top/(c^\top Qc) = A(I - \sigma xx^\top)A^\top$ where x is a normalized vector.
  - Using Fact 2 we can get an orthonormal matrix $U = \begin{bmatrix} c & v_2 & \dots & v_n \end{bmatrix}$.
  - $I - xx^\top = UDU^\top$ where D is the identity matrix with first entry $1 - \sigma$.
  - Then $A(I - \sigma xx^\top)A^\top = BB^\top$ where $B = AUD^{1/2}$ (we know $\sigma < 1$).

Applying an FC layer with weight matrix W to the ellipsoid gives an ellipsoid with matrix $WQW^\top$, so we update the factor as WA. Similarly, convolutions can be applied directly to the A matrix.

Conceivably, the entire forward pass can be done with just vector-vector multiplications, except for the convolutions and FC layers. If A is just being updated to $A^\top(I + xy^\top)$, we can store a chain e.g.

$$A_{out}^\top = A(I + x_0 y_0^\top)(I + x_1 y_1^\top)W \dots (I + x_k y_k^\top),$$

where W is a convolution or FC weight matrix. We maintain the chain in uncomputed form until we compute the loss after the last layer, at which point we collapse the chain. The loss is $A_{out}^{\top}c$ for some vector c, so we first compute $y_k^{\top}c$, then $c + x_k(y_k^{\top}c)$, etc, occasionally applying convolutions or FC layers.

7 Facts

Fact 1.

$$A\mathcal{E}(q, Q) + b = \mathcal{E}(Aq + b, AQA^{\top})$$

Fact 2. *The orthogonal projection of x onto hyperplane $H = \{x : c^{\top}x = 0\}$, expressed in terms of an orthonormal basis for H, is given by $P^{\top}x$ where*

$$P = \left(I - \frac{2}{\|a\|^2}aa^{\top}\right)_{[:,2:]}$$

and $a = \frac{c}{\|c\|} - e_1$. Also the first column which is ommited from the above matrix is c.

Fact 3. *The orthogonal projection of x onto hyperplane $H = \{x : c^{\top}x = 0\}$, expressed in terms of the original basis, is given by $PP^{\top}x$.*

Fact 4. *The orthogonal projection of x onto hyperplane $c^{\top}x = \gamma$, expressed in terms of the original basis, is given by*

$$f(x) = PP^{\top}(x - \gamma c) + \gamma c.$$

Fact 5. *Suppose the line $\overline{pq}$ is parallel to the hyperplane $HP = \{x : x^{\top}c = 0\}$. The sheer matrix that maps p to q while leaving HP fixed is*

$$S = I + c(q - p)^{\top}/(p^{\top}c).$$

(VanArsdale via `http://www.silcom.com/~barnowl/HTransf.htm`)

Fact 6 ([1])**.** *If $-1/n \leq \alpha < 1$, then the minimum volume ellipsoid covering the intersection of $\mathcal{E}(0, I)$ with $H = \{x : e_1^{\top}x \leq -\alpha\}$ is*

$$\mathcal{E}(-\tau e_1,\ \delta(I - \sigma e_1 e_1^{\top}))$$

where

$$\begin{aligned}
\tau &= (1 + n\alpha)/(n + 1), \\
\sigma &= 2(1 + n\alpha)/((n + 1)(n + \alpha)), \\
\delta &= (n^2/(n^2 - 1))(1 - \alpha^2).
\end{aligned}$$

Fact 7 ([1])**.** *Let $\alpha = (c^{\top}q - \gamma)/\sqrt{c^{\top}Qc}$. If $-1/n \leq \alpha < 1$ then the minimum volume ellipsoid covering the intersection of $\mathcal{E}(q, Q)$ with $H = \{x : c^{\top}x \leq \gamma\}$ is*

$$\mathcal{E}(q - \tau(Qc/\sqrt{c^{\top}Qc}),\ \delta(Q - \sigma Qc(Qc)^{\top}/(c^{\top}Qc)))$$

where τ, δ, σ are as above with $\alpha = (c^{\top}q - \gamma)/\sqrt{c^{\top}Qc}$.

Fact 8 ([1]). *If* $-1/n \le \alpha \le -1/n$ *then the minimum volume ellipsoid covering the intersection of* $E = \mathcal{E}(q, Q)$ *with* $H = \{x : c^\top x \le \gamma\}$ *is* E *itself.*

Fact 9. *Affine transformations map lines to lines, and parallel lines to parallel lines.*

Fact 10. *If* f *is affine, and* $\|f(x) - f(y)\| < \|x - y\|$*, then for all* x', y' *on the line* $\overline{xy}$ *we have* $\|f(x') - f(y')\| < \|x' - y'\|$*.*

Fact 11. *The maximum of* $c^\top y$ *over* $\mathcal{E}(q, Q)$ *is*

$$c^\top q + c^\top Qc/\sqrt{c^\top Qc}.$$

(https://math.stackexchange.com/questions/1832467/maximizing-a-linear-function-over-an-ellipsoid)

Fact 12. *The principle axes of* $\mathcal{E}(q, Q)$ *are the eigenvectors of* Q *and the lengths of the semiaxes are the square roots of corresponding eigenvalues.*

References

[1] Robert G Bland, Donald Goldfarb, and Michael J Todd. The ellipsoid method: A survey. *Operations research*, 29(6):1039–1091, 1981.