Principle of Least Squares

We consider two jointly distributed random vectors, x and y. x is n-dimensional and y is m-dimensional. The least square estimation problem is stated as follows. Given the information that y takes the value y_1, what is the best estimate $\hat{x}$ of the corresponding value of the random vector x in the sense of least squares: i.e. find $\hat{x}$ which minimizes

$$E[\|x-\hat{x}\|^2 \,|\, y=y_1] = E[(x-\hat{x})^T(x-\hat{x}) \,|\, y=y_1] \qquad \text{(LS-1)}$$

Theorem: The solution to this problem is given for all y_1 with $p_y(y_1) > 0$ by the conditional expectation,

$$\hat{x} = E[x \,|\, y=y_1] = \int_{-\infty}^{\infty} x p_{x|y}(x|y_1)dx \qquad \text{(LS-2)}$$

The corresponding minimum mean square error is the trace of the conditional covariance of x given y = y_1,

$$trE[(x-\hat{x})(x-\hat{x})^T \,|\, y=y_1] \qquad \text{(LS-3)}$$

where tr denotes the trace of a square matrix.

Proof: Let z be a deterministic n-dimensional vector. Then,

$$\begin{aligned} E[\|x-z\|^2 \,|\, y=y_1] &= E[xx^T|y=y_1] - 2z^T E[x|y=y_1] + z^T z \\ &= E[x^T x|y=y_1] - 2z^T\hat{x} + z^T z + \hat{x}^T\hat{x} - \hat{x}^T\hat{x} \\ &= E[x^T x|y=y_1] + \|z-\hat{x}\|^2 - \hat{x}^T\hat{x} \end{aligned} \qquad \text{(LS-4)}$$

Notice that the first and third terms in the right hand side of Eq. (LS-4) do not depend on z and the second term is non-negative. Clearly,

$$\min_z \{E[\|x-z\|^2 \,|\, y=y_1]\}$$

is achieved for z = $\hat{x}$. With this choice of z,

$$E[\|x-\hat{x}\|^2 |y=y_1] = E[(x-\hat{x})^T(x-\hat{x}) \,|\, y=y_1] = E[tr(x-\hat{x})(x-\hat{x})^T|y=y_1] = trE[(x-\hat{x})(x-\hat{x})^T|y=y_1]$$

(End of Proof)

The least square estimate (conditional expectation) can be determined for every value of y. Therefore, it is a function of the random vector y. When the conditional expectation is regarded as a function of y, it is called the least square estimator and it written as

$$\hat{x} = E[x|y] \tag{LS-5}$$

Notice that till here we have not mentioned the distribution functions of x and y. In particular, Eq. (LS-2) is true regardless of the shapes of probability distribution (density) functions.

When x and y are Gaussian distributed, noting (PR-25) the least square estimator is given by

$$\hat{x}=E[x]+X_{xy}X_{yy}^{-1}(y-E[y]) \tag{LS-6}$$

Notice that Eq. (LS-6) is a linear function of y, which is Gaussian distributed. Therefore, $\hat{x}$ is Gaussian with

$$\begin{aligned}
\textit{mean:}\quad & E[\hat{x}] = E\{E[x]+X_{xy}X_{yy}^{-1}(y-E[y])\} = E[x] \\
\textit{and} \\
\textit{covariance:}\quad & E[(\hat{x}-E[x])(\hat{x}-E[x])^T] \\
& = X_{xy}X_{yy}^{-1}E[(y-E[y])(y-E[y])^T]X_{yy}^{-1}X_{xy}^T \\
& = X_{xy}X_{yy}^{-1}X_{xy}^T
\end{aligned} \tag{LS-7}$$

The least squares estimation error $\tilde{x} = x - \hat{x}$ is a linear function of y. *The mean of $\tilde{x}$ is zero and the covariance is equal to the conditional covariance given by (PR-25)*: i.e. the estimation error covariance is

$$E[\tilde{x}\tilde{x}^T] = X_{xx}-X_{xy}X_{yy}^{-1}X_{yx} \tag{LS-8}$$

Notice that the covariance of the estimation error is smaller than the covariance of x, which means the reduction of uncertainties.

Properties of Least Square Estimate (Gaussian Case)

We have seen that the least square estimate is given by the conditional expectation. The conditional expectation has several important properties when x and y are jointly Gaussian distributed.

Property i. The estimation error $\tilde{x} = x - \hat{x}$ is uncorrelated with y. Furthermore, $\tilde{x}$ and $\hat{x}$ are orthogonal in the sense that

$$E[(x-\hat{x})^T\hat{x}]=0 \tag{LS-9}$$

Proof: Noting that the estimation error is zero mean, the (cross) covariance matrix of $\tilde{x}$ and y is

$$E[(x-\hat{x})(y-E(y))^T] = E[\{x-E[x]-X_{xy}X_{yy}^{-1}(y-E[y])\}(y-E[y])^T] \quad \text{(LS-10)}$$
$$= X_{xy}-X_{xy}X_{yy}^{-1}X_{yy} = 0$$

This implies that the estimation error and y are uncorrelated. The orthogonality of $\tilde{x}$ and $\hat{x}$ are proved as shown below.

$$E[(x-\hat{x})^T\hat{x}] = E[(x-\hat{x})^T\{E[x]+X_{xy}X_{yy}^{-1}(y-E[y])\}]$$
$$= E[x-\hat{x}]^TE[x]+E[(x-\hat{x})^TX_{xy}X_{yy}^{-1}(y-E[y])]$$
$$= tr\{X_{xy}X_{yy}^{-1}E[(y-E[y])(x-\hat{x})^T]\} = 0$$

where we have noted that y and x - $\hat{x}$ are uncorrelated. (End of proof.)

Geometric interpretation of orthogonality is given in the figure shown below.

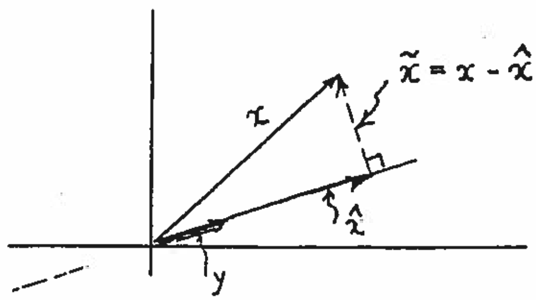


Fig. 1 Geometric Interpretation of Property i.

Property ii. Let y and z be Gaussian and uncorrelated: i.e. $E[(y - E[y])(z - E[z])^T] = 0$. Then,

$$E[x|y,z] = E[x]+(E[x|y]-E[x])+(E[x|z]-E[x]) \quad \text{(LS-11)}$$
$$= E[x|y]+E[x|z]-E[x] = E[x|y]+E[\tilde{x}_{|y}|z]$$

where $\tilde{x}_{|y} = x - E[x|y]$. Note that x can be written as $x = \hat{x}_{|y} + \tilde{x}_{|y}$ where $\hat{x}_{|y} = E[x|y]$. The estimation error covariance for $E[x|y,z]$ is

$$X_{xx}-X_{xy}X_{yy}^{-1}X_{yx}-X_{xz}X_{zz}^{-1}X_{zx} = X_{\tilde{x}\tilde{x}}-X_{xz}X_{zz}^{-1}X_{zx} = X_{\tilde{x}\tilde{x}}-X_{\tilde{x}z}X_{zz}^{-1}X_{z\tilde{x}} \quad \text{(LS-12)}$$

where $X_{\tilde{x}\tilde{x}}=E[\tilde{x}_{|y}\tilde{x}_{|y}^T] = X_{xx}-X_{xy}X_{yy}^{-1}X_{yx}$ is the covariance of $x-E[x|y]$ and $X_{\tilde{x}z} = E[\tilde{x}_{|y}(z - E[z])^T]$.

Proof: Define

$$w = \begin{bmatrix} y \\ z \end{bmatrix}$$

Then

$$E[x|w] = E[x]+X_{xw}X_{ww}^{-1}(w-E[w]) = E[x]+[X_{xy}\ \ X_{xz}]\begin{bmatrix}X_{yy} & X_{yz}\\ X_{zy} & X_{zz}\end{bmatrix}^{-1}(w-E[w])$$
$$= E[x]+X_{xy}X_{yy}^{-1}(y-E[y])+X_{xz}X_{zz}^{-1}(z-E[z]) = E[x]+(E[x|y]-E[x])+(E[x|z]-E[x])$$
$$= E[x|y]+E[\hat{x}_{|y}+\tilde{x}_{|y}|z]-E[x] = E[x|y]+E[\tilde{x}_{|y}|z]$$

where we have noted $X_{yz} = X_{zy}^T = 0$ and $E[\hat{x}_{|y}|z] = E[x]+X_{xy}X_{yy}^{-1}X_{yz}X_{zz}^{-1}(z-E[z]) = E[x]$.

The estimation error covariance is

$$X_{xx}-X_{xw}X_{ww}^{-1}X_{wx} = X_{xx}-X_{xy}X_{yy}^{-1}X_{yx}-X_{xz}X_{zz}^{-1}X_{zx}$$

Noting $E[(y-E[y])(z-E[z])^T]=0$, $X_{xz}(=X_{zx}^T)$ in the third term can be rewritten as shown below.

$$X_{xz} = E[(x-E[x])(z-E[z])^T] = E[(\hat{x}_{|y}+\tilde{x}_{|y}-E[x])(z-E[z])^T]$$
$$= E[(X_{xy}X_{yy}^{-1}(y-E[y])+\tilde{x}_{|y})(z-E[z])^T] = E[\tilde{x}_{|y}(z-E[z])] = X_{\tilde{x}z}$$

(End of proof.)

Property iii. If y and z are Gaussian and correlated,

$$\begin{aligned} E[x|y,z] &= E[x|y,\tilde{z}_{|y}] \\ &= E[x|y]+E[x|\tilde{z}_{|y}]-E[x] \\ &= E[x|y]+E[\tilde{x}_{|y}|\tilde{z}_{|y}] \end{aligned} \tag{LS-13}$$

where

$$\tilde{z}_{|y} = z-E[z|y] \quad and \quad \tilde{x}_{|y} = x-E[x|y]. \tag{LS-14}$$

The covariance of the estimation error is

$$X_{\tilde{x}\tilde{x}}-X_{\tilde{x}\tilde{z}}X_{\tilde{z}\tilde{z}}^{-1}X_{\tilde{z}\tilde{x}} \tag{LS-15}$$

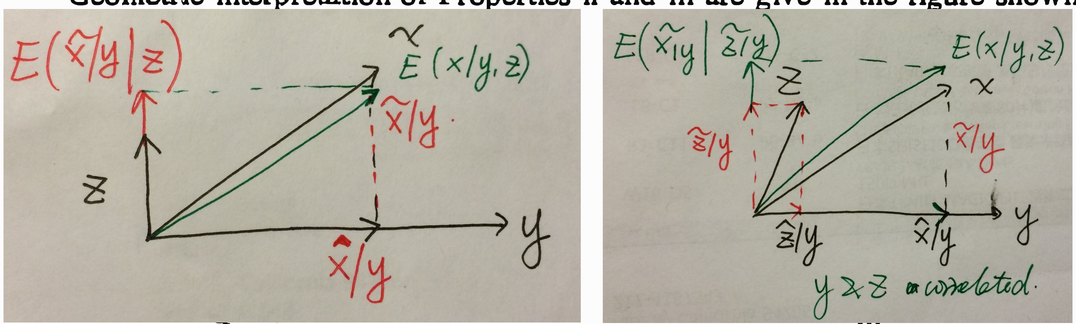
where

$$X_{\tilde{x}\tilde{x}} = E[\tilde{x}_{|y}\tilde{x}_{|y}^T],\ \ X_{\tilde{x}\tilde{z}} = E[\tilde{x}_{|y}\tilde{z}_{|y}^T] \ \ and \ \ X_{\tilde{z}\tilde{z}} = E[\tilde{z}_{|y}\tilde{z}_{|y}^T] \tag{LS-16}$$

Proof: $E[x|y,z] = E[x|y, \tilde{z}_{|y}]$ since a knowledge of y and z is clearly equivalent to a knowledge of y and $\tilde{z}_{|y}$. $\tilde{z}_{|y}$ depends on both y and z. From Property i, y and $\tilde{z}_{|y} = z-E[z|y]$ are uncorrelated. From Property ii, the expressions for $E[x|y,z]$ and the estimation error covariance follow.

(End of proof.)

Geometric interpretation of Properties ii and iii are give in the figure shown below.



Property ii Property iii

Fig. 2 Geometric Interpretation of Properties ii and iii

Example of Least Square Estimation

We consider a problem of estimating the steady state motor velocity based on tachometer measurements. Imagine that a motor is picked up from a lot and that it is placed at a test stand to measure the steady state velocity x under a constant voltage excitation. x is random due to uncertainties in bearing friction and motor constant. Its mean and variance are $E[x] = 10$ rad/s and $E[(x-10)^2] = 2$ rad^2/s^2. Measurement equations are

$$\textit{First tachometer:} \quad y_1 = x+v_1$$
$$\textit{Sencond tachometer:} \quad y_2 = x+v_2$$

where measurement noises satisfy

$$E[v_1] = E[v_2] = 0,\ E[v_1^2] = E[v_2^2] = 1\ rad^2/s^2,\ E[v_1 v_2] = 0,\ E[(x-10)v_i] = 0$$

1) Best estimate of x in terms of y_1

Notice the following covariances.

$$\begin{aligned} X_{xy_1} &= E[(x-E[x])(y_1-E[y_1])] \\ &= E[(x-E[x])(x+v_1-E[x]-E[v_1])] = E[(x-E[x])^2]+E[(x-E[x])v_1] = 2 \\ X_{y_1y_1} &= E[(y_1-E[y_1])^2] = E[(x+v_1-E[x]-E[v_1])^2] \\ &= E[(x-E[x])^2]+2E[(x-E[x])v_1]+E[v_1^2] = 3 \end{aligned}$$

Therefore,

$$E[x|y_1] = E[x]+X_{xy_1}X_{y_1y_1}^{-1}(y_1-E[y_1]) = 10+\frac{2}{3}(y_1-10)$$

2) Best estimate of x in terms of y_2

Exercise.

3) Best estimate of x in terms of y_1 and y_2

First approach: Define $y = [y_1 \ \ y_2]^T$. Then,

$$X_{xy} = E[(x-E[x])\begin{bmatrix} y_1-E[y_1] \\ y_2-E[y_2] \end{bmatrix}^T] = [2 \ \ 2]$$

$$X_{yy} = E[\begin{bmatrix} y_1-E[y_1] \\ y_2-E[y_2] \end{bmatrix}[y_1-E[y_1] \ \ y_2-E[y_2]]] = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

Therefore,

$$E[x|y] = E[x]+X_{xy}X_{yy}^{-1}(y-E[y]) = 10+[2 \ \ 2]\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1}\begin{bmatrix} y_1-10 \\ y_2-10 \end{bmatrix}$$

$$= 10+\frac{2}{5}(y_1-10)+\frac{2}{5}(y_2-10)$$

Second approach (use property iii):

From Property iii,

$$E[x|y_1,y_2] = E[x|y_1]+E[\tilde{x}_{|y_1}|\tilde{y}_{2|y_1}]$$

where

$$\tilde{x}_{|y_1} = x-E[x|y_1] = x-[10+\frac{2}{3}(y_1-10)] = \frac{1}{3}(x-10)-\frac{2}{3}v_1$$

$$\tilde{y}_{2|y_1} = y_2-\hat{y}_{2|y_1} = y_2-[E[y_2]+X_{y_2y_1}X_{y_1y_1}^{-1}(y_1-E[y_1])$$

$$= y_2-[10+\frac{2}{3}(y_1-10)] = (y_2-10)-\frac{2}{3}(y_1-10)$$

The following quantities can easily be calculated.

$$E[\tilde{x}_{|y_1}] = 0, \ E[\tilde{y}_{2|y_1}] = 0, \ X_{\tilde{x}_{|y_1}\tilde{y}_{2|y_1}} = \frac{2}{3} \ and \ X_{\tilde{y}_{2|y_1}\tilde{y}_{2|y_1}} = \frac{5}{3}$$

Therefore,

$$E[x|y_1,y_2] = E[x|y_1]+E[\tilde{x}_{|y_1}]+X_{\tilde{x}_{|y_1}\tilde{y}_{2|y_1}}X_{\tilde{y}_{2|y_1}\tilde{y}_{2|y_1}}^{-1}(\tilde{y}_{2|y_1}-E[\tilde{y}_{2|y_1}])$$

$$= 10+\frac{2}{3}(y_1-10)+\frac{2}{5}[(y_2-10)-\frac{2}{3}(y_1-10)] = 10+\frac{2}{5}(y_1-10)+\frac{2}{5}(y_2-10)$$

Example of Least Squares (Gaussian Case)

Note. This example is a generalized version of one in Reader (page LS-5)

$$y_i = x + v_i \qquad i = 1, 2$$

$$E[x] = x_0, \quad E[(x-x_0)^2] = X_0, \quad E[v_i] = 0, \quad E[v_i v_j] = V\delta_{ij}$$

$$E[(x-x_0)v_i] = 0 \qquad \left(\begin{array}{l}\delta_{ij} = 1 \text{ for } i=j=1 \text{ or } 2 \\ \phantom{\delta_{ij}} = 0 \text{ for } i \neq j\end{array}\right)$$

1) $i = 1$ Best estimate of x based on y_1.

$$E[x|y_1] = E[x] + X_{xy_1} X_{y_1 y_1}^{-1} [y_1 - E[y_1]]$$

Computation of quantities in the above expression

$$E[x] = x_0, \quad E[y_1] = E[x + v_1] = x_0$$

$$X_{xy_1} = E[(x-x_0)(\underbrace{x+v_1}_{y} - x_0)] = E[(x-x_0)^2] + \underbrace{E[(x-x_0)v_1]}_{0} = X_0$$

$$X_{y_1 y_1} = E[(x+v_1-x_0)(x+v_1-x_0)]$$

$$= E[(x-x_0)^2] + 2\underbrace{E[(x-x_0)v_1]}_{0} + E[v_1^2] = X_0 + V$$

$$\Rightarrow E[x|y_1] = x_0 + \frac{X_0}{X_0+V}[y_1 - x_0] = \frac{V}{X_0+V}x_0 + \frac{X_0}{X_0+V}y_1$$

The best estimate is a weighted sum of x_0 and y_1. The two weighting factors depend on V and X_0. For example, if $V \ll X_0$, then $\frac{V}{X_0+V} \ll 1$ and $\frac{X_0}{X_0+V} \approx 1$.

The estimation error covariance is

$$X_{xx} - X_{xy_1} X_{y_1 y_1}^{-1} X_{y_1 x} = X_0 - \frac{X_0^2}{X_0+V} = \frac{X_0 V}{X_0+V}$$

2) Best estimate of x based on y_2. (Essentially the same as 1).)

$$E[x|y_2] = E[x] + X_{xy_2} X_{y_2 y_2}^{-1} [y_2 - E[y_2]]$$

$$= x_0 + \frac{X_0}{X_0+V}[y_2 - x_0]$$

3) Best estimate of x based on y_1 and y_2:

First approach: $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

$$E[x|y_1, y_2] = E[x] + X_{xy} X_{yy}^{-1} \begin{bmatrix} y_1 - E[y_1] \\ y_2 - E[y_2] \end{bmatrix}$$

$$X_{xy} = E\left[(x - x_0)\begin{bmatrix} x + v_1 - x_0 \\ x + v_2 - x_0 \end{bmatrix}^T\right] = [X_0 \;\; X_0]$$

$$X_{yy} = E\left[\begin{bmatrix} x + v_1 - x_0 \\ x + v_2 - x_0 \end{bmatrix}[x + v_1 - x_0, x + v_2 - x_0]\right] = \begin{bmatrix} X_0 + V & X_0 \\ X_0 & X_0 + V \end{bmatrix}$$

$$\Rightarrow E[x|y_1, y_2] = x_0 + [X_0, X_0]\begin{bmatrix} X_0 + V & X_0 \\ X_0 & X_0 + V \end{bmatrix}^{-1}\begin{bmatrix} y_1 - x_0 \\ y_2 - x_0 \end{bmatrix}$$

$$\frac{1}{2X_0V + V^2}\begin{bmatrix} X_0 + V & -X_0 \\ -X_0 & X_0 + V \end{bmatrix}$$

$$\Rightarrow E[x|y_1, y_2] = x_0 + \frac{X_0}{2X_0 + V}[y_1 - x_0] + \frac{X_0}{2X_0 + V}[y_2 - x_0]$$

$$= \frac{V}{2X_0 + V}x_0 + \frac{X_0}{2X_0 + V}y_1 + \frac{X_0}{2X_0 + V}y_2$$

The estimation error covariance is

$$X_{xx} - X_{xy}X_{yy}^{-1}X_{yx} = X_0 - \frac{1}{2X_0V + V^2}[X_0 \; X_0]\begin{bmatrix} X_0 + V & -X_0 \\ -X_0 & X_0 + V \end{bmatrix}\begin{bmatrix} X_0 \\ X_0 \end{bmatrix}$$

X_0

$$= X_0 - \frac{2X_0^2}{2X_0 + V}$$

Second approach: (Use Property c).

$$E[x|y_1, y_2] = E[x|y_1] + E[\tilde{x}_{|y_1} | \tilde{y}_{2|y_1}]$$

$$\tilde{x}_{|y_1} = x - E[x|y_1] = x - \left\{x_0 + \frac{X_0}{X_0 + V}[\underbrace{y_1}_{x + v_1} - x_0]\right\}$$

$$= \frac{V}{X_0 + V}(x - x_0) - \frac{X_0}{X_0 + V}v_1$$

$$\tilde{y}_{2|y_1} = y_2 - E[y_2|y_1] = x + v_2 - \left\{x_0 + \overset{X_0}{X_{y_2y_1}} \overset{1/X_0+V}{X_{y_1y_1}^{-1}}[y_1 - \overset{E[y_1]}{x_0}]\right\}$$

$$= \frac{V}{X_0 + V}(x - x_0) + v_2 - \frac{X_0}{X_0 + V}v_1$$

$$\Rightarrow \quad E[\tilde{x}_{|y_1} \mid \tilde{y}_{2|y_1}] = \cancelto{0}{E[\tilde{x}_{|y_1}]} + X_{\tilde{x}_{|y_1}\tilde{y}_{2|y_1}} X^{-1}_{\tilde{y}_{2|y_1}\tilde{y}_{2|y_1}} \left[\tilde{y}_{2|y_1} - \cancelto{0}{E[\tilde{y}_{2|y_1}]}\right]$$

$$X_{\tilde{x}_{|y_1}\tilde{y}_{2|y_1}} = E\left[\left(\frac{V}{X_0+V}(x-x_0) - \frac{X_0}{X_0+V}v_1\right)\left(\frac{V}{X_0+V}(x-x_0) + v_2 - \frac{X_0}{X_0+V}v_1\right)\right]$$

$$= \frac{V^2}{(X_0+V)^2}X_0 + \frac{X_0^2}{(X_0+V)^2}V = \frac{X_0 V}{X_0+V}$$

$$X_{\tilde{y}_{2|y_1}\tilde{y}_{2|y_1}} = E\left[\left(\frac{V}{X_0+V}(x-x_0) + v_2 - \frac{X_0}{X_0+V}v_1\right)^2\right]$$

$$= \frac{V^2}{(X_0+V)^2}X_0 + V + \frac{X_0^2}{(X_0+V)^2}V = \frac{2VX_0+V^2}{X_0+V}$$

$$= \frac{V^2X_0 + V(X_0+V)^2 + X_0^2V}{(X_0+V)^2} \qquad X_0V(X_0+V)$$

$$\Rightarrow \quad E[\tilde{x}_{|y_1} \mid \tilde{y}_{2|y_1}] = \frac{\cancel{X_0V}}{\cancel{X_0+V}} \cdot \frac{\cancel{X_0+V}}{2VX_0+V^2}\Bigg[\underbrace{y_2 - x_0 - \frac{X_0}{X_0+V}(y_1-x_0)}_{\hat{y}_{2|y_1}}\Bigg]$$

$$= \frac{X_0}{2X_0+V}\left[(y_2-x_0) - \frac{X_0}{X_0+V}(y_1-x_0)\right]$$

$$\Rightarrow \quad E[x \mid y_1, y_2]$$

$$= x_0 + \frac{X_0}{X_0+V}[y_1-x_0] + \frac{X_0}{2X_0+V}\left[(y_2-x_0) - \frac{X_0}{X_0+V}(y_1-x_0)\right]$$

$$= x_0 + \frac{X_0}{2X_0+V}[y_1-x_0] + \frac{X_0}{2X_0+V}[y_2-x_0]$$

$$= \frac{V}{2X_0+V}x_0 + \frac{X_0}{2X_0+V}y_1 + \frac{X_0}{2X_0+V}y_2$$