Inverse problems

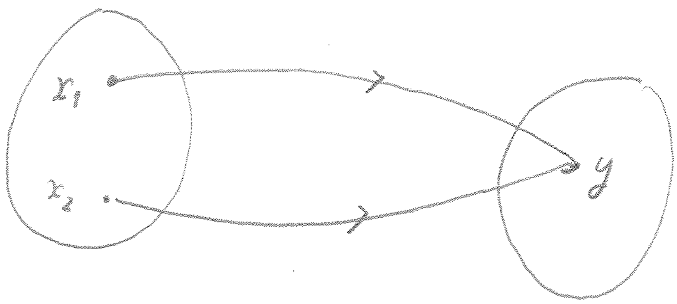
Inverse problem: Retrieve information of an unknown quality by

(a) indirect observation

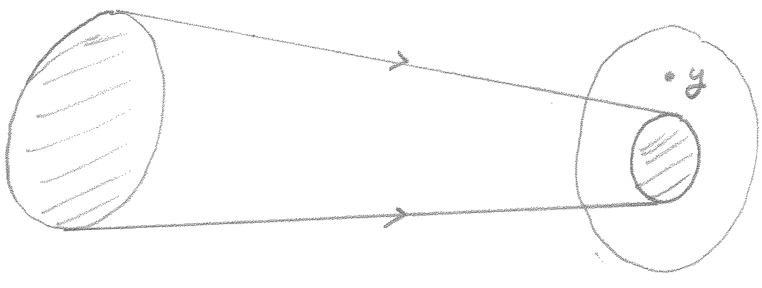
(b) additional information of the unknown.

Ill-posedness:

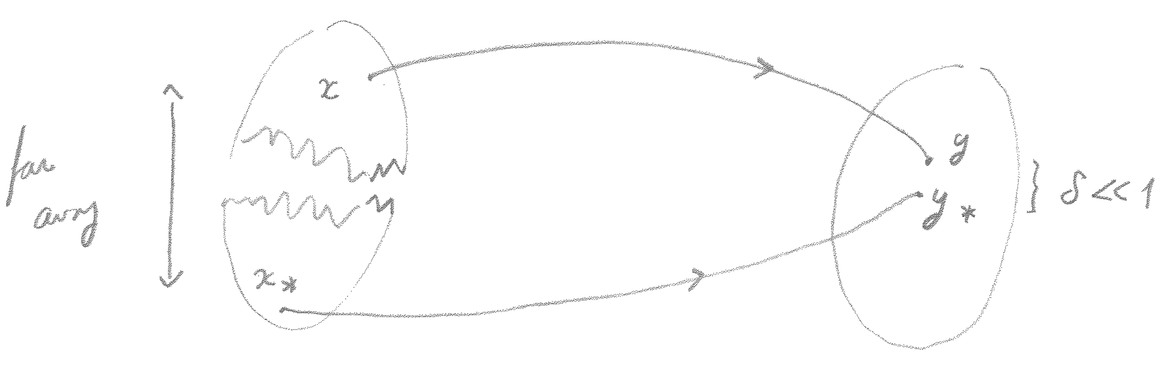
1: Several feasible solutions:



2: No solution:



3: Ill-posedness:



1. Linear algebra primer, SVD

Matrices and vectors:

$$A \in \mathbb{R}^{m\times n}, \quad A = \begin{bmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & & \vdots \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix}$$

$$x \in \mathbb{R}^n, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Matrix as a collection of column vectors:

$$A = [a_1, a_2, \dots, a_n], \quad a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

Matrix-vector product

$$Ax = b, \quad b_j = \sum_{k=1}^{n} a_{jk} x_k$$

$$b = [a_1, a_2, \dots, a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{k=1}^{n} x_k a_k$$

= linear combination of the columns of A with coefficients x_j, $1 \le j \le n$.

Let $v_1, v_2, \dots, v_n \in \mathbb{R}^m$

$$sp\{v_1, v_2, \dots, v_n\} = \{b \mid b = \sum_{j=1}^{n} \alpha_j v_j\}$$

<u>Range</u> of A

$$R(A) = sp\{a_1, a_2, \dots, a_n\}$$
$$= \{Ax \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^m$$

<u>Null space</u> of A

$$N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} \subset \mathbb{R}^n$$

Vectors $v_1, \dots v_n \in \mathbb{R}^m$ are <u>linearly independent</u> if

$$\sum_{j=1}^{n} \alpha_j v_j = 0 \iff \alpha_1 = \alpha_2 = \dots = \alpha_n = 0$$

Linearly independent vectors $v_1, \dots, v_n$ span an <u>n-dimensional</u> space

Vectors $v_1, \dots v_n \in \mathbb{R}^m$ can be linearly independent only if $n \le m$.

Matrix $A \in \mathbb{R}^{m \times n}$ is of <u>full column rank</u> if $A = [a_1, \dots, a_m]$ and

$$\dim(sp\{a_1, \dots a_m\}) = k = \min(n, m),$$

i.e. the columns span a maximal possible subspace

<u>Transpose</u> of a matrix $A \in \mathbb{R}^{m \times n}$

$$A^T = \begin{bmatrix} a_{11} & a_{21} \cdots & a_{m1} \\ \vdots & & \\ a_{1n} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \in \mathbb{R}^{n \times m}$$

$$a_j^T = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}^T = [a_{1j} \ \ a_{2j} \ \cdots \ a_{mj}]$$

Transpose of a matrix-vector product

$$(Ax)^T = x^T A^T$$

Matrix product:

$$A \in \mathbb{R}^{m \times n}, \ B = \mathbb{R}^{n \times k} = [b_1, \ldots, b_k]$$

$$AB = [\underbrace{Ab_1}_{\in \mathbb{R}^m} \ Ab_2 \ \cdots \ Ab_k] \in \mathbb{R}^{m \times k}$$

$$(AB)^T = \begin{bmatrix} (Ab_1)^T \\ \vdots \\ (Ab_k)^T \end{bmatrix} = \begin{bmatrix} b_1^T A^T \\ \vdots \\ b_k^T A^T \end{bmatrix} = \begin{bmatrix} b_1^T \\ \vdots \\ b_k^T \end{bmatrix} A^T = B^T A^T$$

<u>Inner product</u> of vectors:

$$x, y \in \mathbb{R}^n; \qquad x^T y = \sum_{j=1}^{n} x_j y_j$$

$$= [x_1, \dots, x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

<u>Outer product</u> of vectors

$$x \in \mathbb{R}^m, y \in \mathbb{R}^n;$$

$$x y^T = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} [y_1, \dots y_n] = \begin{bmatrix} x_1 y_1 & x_1 y_2 \dots & x_1 y_n \\ \vdots & & \\ x_m y_1 & \dots & x_m y_n \end{bmatrix}$$

$$\in \mathbb{R}^{m \times n}$$

<u>Observe</u>: if $\alpha \in \mathbb{R}^n$,

$$(x y^T)\alpha = x \underbrace{(y^T \alpha)}_{\in \mathbb{R}} \in \operatorname{sp}[x]$$

i.e., $x y^T$ is a rank-1 matrix.

More generally, if

$$u_1, \dots, u_k \in \mathbb{R}^m, \quad v_1, \dots, v_k \in \mathbb{R}^n,$$

$$A = \sum_{j=1}^{k} u_j v_k^T \in \mathbb{R}^{m \times n}$$

is <u>at most</u> rank-k matrix:

$$Ax = \sum_{j=1}^{k} u_j \underbrace{(v_k^T x)}_{\in \mathbb{R}} \in \mathrm{sp}\{u_1, \ldots, u_k\}$$

We may write

$$A = UV^T, \quad U = [u_1, \ldots u_k] \in \mathbb{R}^{n \times k},$$

$$V = [v_1, \ldots v_k] \in \mathbb{R}^{m \times k}.$$

Vectors $u, v \in \mathbb{R}^n$ are <u>orthogonal</u>, if

$$u^T v = v^T u = 0,$$

and we write $u \perp v$.

A matrix $U \in \mathbb{R}^{n \times n}$, $U = [u_1, \ldots, u_n]$ is an orthogonal matrix, if

(1) Its column vectors are mutually orthogonal,

$$u_j \perp u_k, \quad j \neq k$$

(2) The columns have <u>unit norm</u>,

$$\|u_j\| =: \sqrt{u_j^T u_j} = 1, \quad 1 \le j \le n.$$

In particular,

$$U^TU = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} [u_1, \dots, u_n] = \begin{bmatrix} \overbrace{u_1^Tu_1}^{=1} & \overbrace{u_1^Tu_2}^{=0} & \dots & u_1^Tu_n \\ u_2^Tu_1 & u_2^Tu_2 & & \\ \vdots & & & \\ u_n^Tu_1 & \dots & \dots & u_n^Tu_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} = I = \text{unit matrix}$$

Orthogonal matrices preserve the vector norm:

$$\|Ux\|^2 = (Ux)^T(Ux) = x^T\underbrace{U^TU}_{I}x = x^Tx$$
$$= \|x\|^2$$

Observe: If $u_1, \dots u_n \in \mathbb{R}^m$ are orthogonal, and $n \leq m$, then

$$\dim(\operatorname{sp}\{u_1, \dots, u_n\}) = n$$

The vectors $\{u_1, \dots, u_n\}$ form an orthogonal basis of the space

$$S = \operatorname{sp}\{u_1, \dots u_n\}.$$

Two subspaces $M, N \subset \mathbb{R}^n$ are orthogonal.

$$M \perp N,$$

if
$$x \in M,\ y \in N \Rightarrow x^T y = 0.$$
The <u>orthogonal complement</u> of the subspace $N \subset \mathbb{R}^n$ is

$$N^\perp = \{ y \in \mathbb{R}^n \mid y \perp x \ \forall\, x \in N \}$$

We have

$$R(A)^\perp = N(A^T)$$
which can be shown as follows:

$$\begin{aligned} y \in N(A)^\perp &\Leftrightarrow y^T(Ax) = 0 \ \forall\, x \in \mathbb{R} \\ &\Leftrightarrow (y^T A)x = 0 \ \forall\, x \in \mathbb{R}^n \\ &\Leftrightarrow (A^T y)^T x = 0 \ \forall\, x \in \mathbb{R}^n \\ &\Leftrightarrow A^T y = 0, \end{aligned}$$

which can be seen, e.g. by choosing $x = A^T y$.

A central concept in the discussion to ensue is the <u>Singular Value Decomposition</u> (SVD) of a matrix:

For every $A \in \mathbb{R}^{n \times m}$ exists orthogonal matrices $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{m \times m}$ and a diagonal matrix $D \in \mathbb{R}^{n \times m}$ with non-negative diagonal entries such that

$$A = U D V^T$$

Before discussing it further, let's see what diagonal matrices look like:

If $m > n$,

$$D = \underbrace{\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \\ \hdashline & & 0 & \end{bmatrix}}_{n} \Bigg\} m = \mathrm{diag}(d_1, \dots, d_n)$$

If $m \leq n$,

$$D = \underbrace{\left[\begin{array}{ccc:c} d_1 & & & \\ & \ddots & & 0 \\ & & d_m & \end{array}\right]}_{n} \Big\} m = \mathrm{diag}(d_1, \dots, d_m)$$

so we may write

$$D = \mathrm{diag}\,(d_1, \ldots, d_{\min\{n,m\}}) \;, \quad \overbrace{d_1 \geq d_2 \geq \ldots d_{\min\{n,m\}}}^{\geq 0}$$

The numbers $d_j \geq 0$ are the <u>singular values</u> of A.

Let us write

$$U = [u_1, u_2, \ldots, u_m] \;, \; V = [v_1, \ldots, v_n] \,,$$

The SVD can be written as

$$A = \sum_{j=1}^{r} d_j u_j v_j^T \,.$$

i.e., A is decomposed into a sum of rank-1 matrices

To show the existence of the SVD decomposition, we need the matrix norm,

$$\|A\| = \max \{ \|Ax\| \mid x \in \mathbb{R}^n, \|x\| = 1 \}$$

Note that since the unit sphere in $\mathbb{R}^n$ is compact, the maximum is indeed reached with some $x \in \mathbb{R}^n$, i.e., there are vectors

$x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\|x\| = \|y\| = 1$, such that

$$Ax = d\,y, \quad d = \|A\| \quad \left(y = \frac{Ax}{\|Ax\|}\right)$$

Let $\{u_2, \dots, u_m\}$, $\{v_2, \dots, v_n\}$ vectors so that

$\{y, u_2, u_3, \dots, u_m\}$ is an orthonormal basis of $\mathbb{R}^m$

$\{x, v_2, \dots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$

Denote $U' = [u_2, \dots, u_m] \in \mathbb{R}^{m \times (m-1)}$

$$V' = [v_2, \dots, v_n] \in \mathbb{R}^{n \times (n-1)}$$

Now,

$$A[x, V'] = [dy, AV']$$

and

$$\begin{bmatrix} y^T \\ U'^T \end{bmatrix} A[x, V'] = \begin{bmatrix} y^T \\ U'^T \end{bmatrix} [dy, AV']$$

$$= \begin{bmatrix} d & y^T(AV') \\ 0 & U'^T(AV') \end{bmatrix} \equiv \begin{bmatrix} d & w^T \\ 0 & B \end{bmatrix} = A_1$$

Now

$$A_1 \begin{bmatrix} d \\ w \end{bmatrix} = \begin{bmatrix} d^2 + w^T w \\ B w \end{bmatrix},$$

so

$$\left\| A_1 \begin{bmatrix} d \\ w \end{bmatrix} \right\| \geq d^2 + \|w\|^2$$

On the other hand

$$\left\| A_1 \begin{bmatrix} d \\ w \end{bmatrix} \right\| \leq \|A_1\| \left\| \begin{bmatrix} d \\ w \end{bmatrix} \right\| = \|A_1\| \sqrt{d^2 + \|w\|^2},$$

so combining,

$$\|A_1\| \geq \sqrt{d^2 + \|w\|^2}$$

But $U = [y, U']$ and $V = [x, V']$ are orthogonal, so

$$\|A_1\| = \|U^T A V\| = \|A\|, \qquad (Ex.)$$

so we must have $w = 0$, i.e.,

$$U A V^T = \begin{bmatrix} d & 0 \\ 0 & B \end{bmatrix}$$

Inductively, we may continue and apply the same reasoning to $B \in \mathbb{R}^{(m-1)\times(n-1)}$.

There are several alternative formulations of SVD: Recall that

$$A = \sum_{j=1}^{r} d_j u_j v_j^T$$

Then

$$A v_k = \sum_{j=1}^{r} d_j u_j \underbrace{v_j^T v_k}_{=\delta_{j,k}} = \begin{cases} d_k u_k, & 1 \le j \le r \\ 0, & k \ge r \end{cases}$$

or, in matrix form

$$AV = UD$$

This gives also the reduced form SVD.

Assume first that $n \le m$

$$A = \begin{bmatrix} x & x & x \\ x & & \\ \vdots & & \\ x & & \end{bmatrix}$$

$$V = \underbrace{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}_{n}\Big\}\, n \quad , \qquad D = m\left\{ \underbrace{\begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \\ \hdashline & 0 & \end{bmatrix}}_{n} \right. \begin{matrix} \big\}\, n \ (D_1) \\ \\ \\ \end{matrix}$$

$$U = \begin{array}{c} \overbrace{\hphantom{U_1}}^{n} \\ \underbrace{\left[\begin{array}{c:c} U_1 & U_2 \end{array}\right]}_{m} \end{array} \Big\}\, m$$

$$AV = D_1 U_1 \quad , \qquad U_2 \perp R(A)$$

Similarly, if $m < n$,

$$A = \underbrace{\begin{bmatrix} \times & & \times \\ \times & & \times \\ \vdots & & \times \\ \times & & \times \end{bmatrix}}_{n}\Big\}\, m$$

$$V = \begin{array}{c} \overbrace{\hphantom{V_1}}^{m} \\ \underbrace{\left[\begin{array}{c:c} V_1 & V_2 \end{array}\right]}_{n} \end{array} \Big\}\, n$$

$$D = \underbrace{\left[\begin{array}{c|c} \overbrace{\begin{matrix} d_1 & & \\ & \ddots & \\ & & d_m \end{matrix}}^{m} & 0 \end{array}\right]}_{n} \Big\}\, m \quad , \quad U = \underbrace{\left[\begin{matrix} \quad & \quad \end{matrix}\right]}_{m} \Big\}\, m$$

$$\begin{cases} A V_1 = D_1 U \\ A V_2 = 0 \end{cases}$$

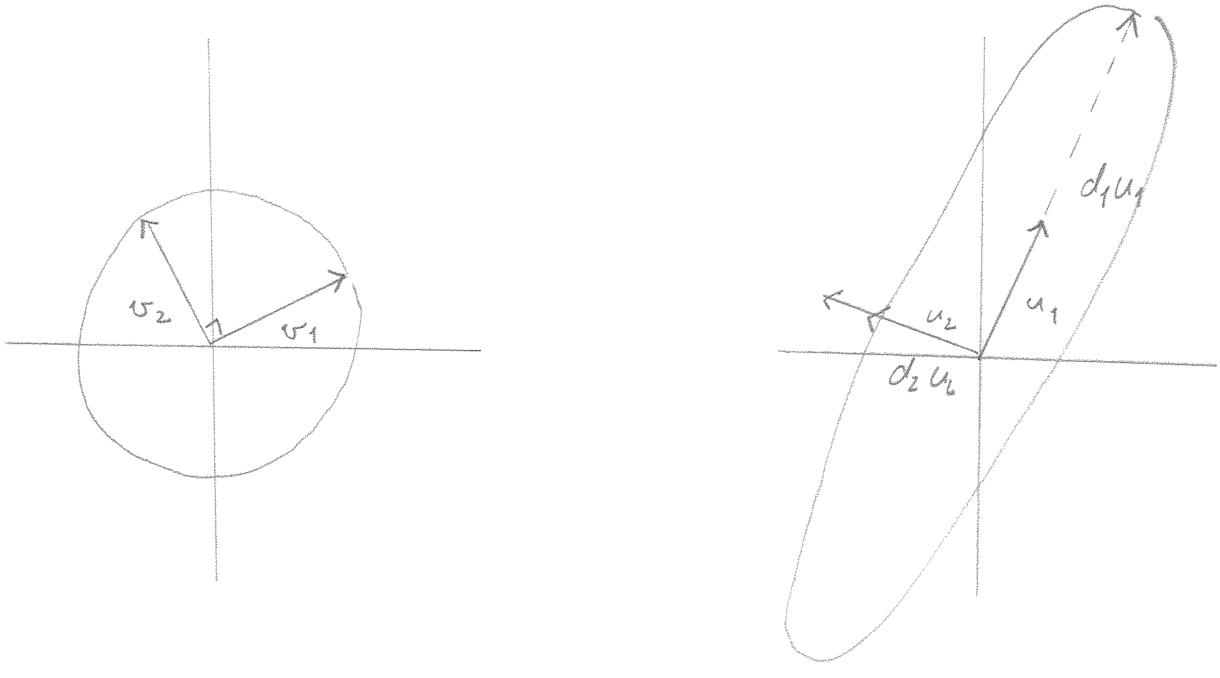
<u>Geometric interpretation</u>: A matrix maps a ball to an ellipsoid:

$$A \in \mathbb{R}^{2 \times 2}$$

<u>Singular system of A</u>:

$$\{ (v_1, u_1, d_1), (v_2, u_2, d_2) \}$$

$$\begin{cases} A v_1 = d_1 u_1 \\ A v_2 = d_2 u_2 \end{cases} \qquad v_1 \perp v_2 , \quad u_1 \perp u_2$$



The ratio $\frac{d_2}{d_1}$ describes the degree of deformation or streching by the matrix.

1.1 SVD and ill-posedness

Consider a linear inverse problem of estimating $x \in \mathbb{R}^n$ from the data

$$y = Ax_* + e \in \mathbb{R}^m \quad , \quad x_* = \text{true } x$$

where $e \in \mathbb{R}^m$ is an error vector. What can go wrong?

1. Null space of A: Suppose we seek to solve the equation

$$y = Ax$$

If $x_0 \in N(A)$, we have

$$A x = A(x + x_0),$$

i.e., if $N(A) \neq \{0\}$, the solution cannot be unique.

Characterization of the null space: Let

$$A = UDV^T,$$

$$D = \mathrm{diag}(d_1, d_2, \ldots, d_k) \quad , k = \min[n, m]$$

Assume that

$$d_1 \geq d_2 \geq \ldots \geq d_p > d_{p+1} = \ldots = d_k = 0$$

Write

$$V = [v_1, \ldots, v_p, v_{p+1}, \ldots, v_n] \in \mathbb{R}^{n \times n}$$

We observe that

$$A v_j = \begin{cases} d_j u_j \neq 0, & \text{if } j \leq p \\ 0 & \text{if } p+1 \leq j \leq n. \end{cases}$$

Conclusion:

$$N(A) = \mathrm{sp}\{v_{p+1}, \ldots, v_n\}$$

<u>2. Range of A:</u> As above, assume that $y \in \mathbb{R}^m$ and for some $x \in \mathbb{R}^n$

$$A x = y$$

Write

$$x = \sum_{j=1}^{n} x_j v_j \quad , \quad x_j = v_j^T x$$

Then

$$A x = \sum_{j=1}^{n} x_j A v_j = \sum_{j=1}^{n} x_j d_j u_j$$

$$= \sum_{j=1}^{p} x_j d_j u_j \quad = y,$$

That is,

$$y \in \mathrm{sp}\{u_1, \dots, u_p\} = R(A)$$

If y has a non-zero component in $R(A)^{\perp}$, that is

$$u_j^T y \neq 0 \quad \text{for some } j = p+1, \dots, m$$

then the equation $y = Ax$ <u>has no</u> <u>solutions</u>.

3. Ill-posedness: Assume that $y \in R(A)$,

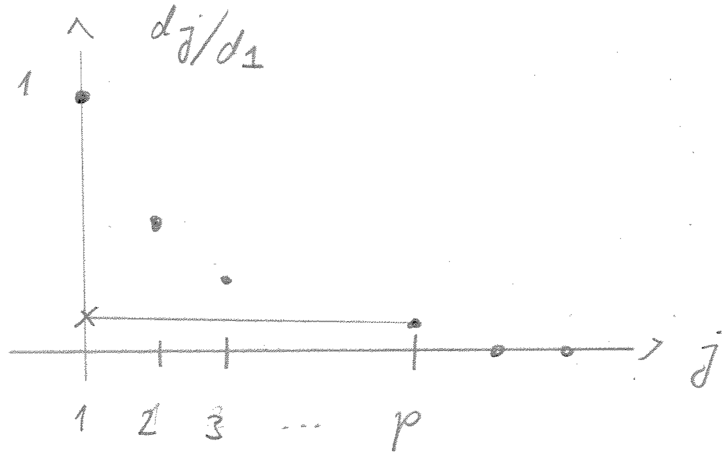
$$y = \sum_{j=1}^{p} y_j u_j$$

The equation $Ax = y$ reads then

$$d_j x_j = y_j, \qquad 1 \le j \le p$$

or

$$x = \sum_{j=1}^{p} \left(\frac{y_j}{d_j}\right) v_j = \frac{1}{d_1} \sum_{j=1}^{p} \left(\frac{d_1}{d_j}\right) y_j v_j$$



If $\frac{d_p}{d_1} \ll 1$, random errors in y are amplified by a factor $\frac{d_1}{d_p} \gg 1$, leading typically to noisy or useless estimates of the vector x.