Chapter 3

Position and Time Determination

The focus of the present chapter is on determining the position of a receiver by measuring ranges or pseudoranges with respect to a set of satellites, whose positions are known. This position is obtained by intersecting a number of spheres, when ranges are measured. Today's global satellite navigation systems use pseudoranges rather than ranges. In this case, the position (and offset of the receiver's clock) is obtained by intersecting a number of coni. The actual calculation of a position and clock offset is performed by a combination of the Newton-Raphson method for finding zeros of a non-linear function, and least-squares estimation. The impact of measurement inaccuracies on the solution is discussed.

3.1 Geometry of Positioning

3.1.1 Position from Ranges

The simplest interesting configuration for determining a position from range measurements is two dimensional. The task is to determine the position $\vec{r}$ of a receiver, given its distance s^1 and s^2 from two distinct satellites[1], see Figure 3.1. The associated equation reads:

$$\|\vec{r} - \vec{r}^{\,1}\| = s^1, \quad \text{and} \quad \|\vec{r} - \vec{r}^{\,2}\| = s^2.$$

Under the assumption that the satellites are not collocated it is possible to indicate two orthogonal vectors that span the two dimensional plane:

$$\vec{e}^{\,12} = \frac{(\vec{r}^{\,2} - \vec{r}^{\,1})}{\|\vec{r}^{\,2} - \vec{r}^{\,1}\|}, \quad \text{and} \quad \vec{n} \perp \vec{e}^{\,12}.$$

The receiver's position can then be represented as a linear combination of these vectors:

$$\vec{r} = \vec{r}^{\,1} + \alpha\vec{e}^{\,12} + \beta\vec{n},$$

with the coefficient α, and β being the projections of $\vec{r} - \vec{r}^{\,1}$ onto $\vec{e}^{\,12}$, and $\vec{n}$. Let $s^{12} = \|\vec{r}^{\,2} - \vec{r}^{\,1}\|$ be the inter-satellite distance, then

$$\|\vec{r} - \vec{r}^{\,1}\|^2 = \alpha^2 + \beta^2 = (s^1)^2,$$

$$\|\vec{r} - \vec{r}^{\,2}\|^2 = \|(\alpha - s^{12})\vec{e}^{\,12} + \beta\vec{n}\|^2 = (\alpha - s^{12})^2 + \beta^2 = (s^2)^2.$$

[1] Upper indices are used for satellites. This is done to ensure consistency with later chapters that involve several satellites and various receivers, such as mobile terminals and reference stations. The ground receivers will be indexed using lower indices. This notation is widely used in the satellite navigation literature. With a little exercise, a confusion with powers is unlikely to occur. The square of r^1 is written as $(r^1)^2$, for example.

It is straight forward to solve these equations:

$$\alpha = \frac{(s^{12})^2 + (s^1)^2 - (s^2)^2}{2\, s^{12}},$$
$$\beta^2 = (s^1)^2 - \alpha^2$$

They are easily generalized to higher dimensions (see Exercise 1). The same symmetry exists in any dimension K. In the non-degenerate case, the satellite might be above or below the $(K-1)$-dimensional subspace spanned by the K satellites. In three dimensional space one solution is typically on earth, while the other one is somewhere in space at twice the distance to the plane spanned by the satellites. The geometry for the three-dimensional case is shown in Figure 3.2.

The wavefront of a signal transmitted by a satellite is a spherical shell $\mathcal{S}$, which grows over time. Let t^k be the transmit time, and t the receive time, then

$$\|\vec{r} - \vec{r}^{\,k}\| = c(t - t^k) = c\tau^k \tag{3.1}$$

A receiver, which can measure τ^k, and thus identify shells, will know that it is at the intersection of particular shells. The intersection of two shells $\mathcal{S}^1$ and $\mathcal{S}^2$ is a circle, and the intersection of the circle with the third shell $\mathcal{S}^3$ is the mentioned degenerate solution with two points.

Any additional satellite that is not in the $n-1$-dimensional space spanned by the n satellites makes the solution unique and over-determined. In the presence of noise a least-squares approach is chosen for finding the best match.

The satellite system "Precise Range and Range Rate Equipment (PRARE)" proposed and developed by Hartl and Reigber [1] performs two-way range measurements for determining the position of the satellite. For this purpose it sends an interrogation signal to the ground that is answered by a set of transponders. PRARE was developed as a satellite tracking and geodetic surveying system. Beidou-1, the Chinese Double Star Positioning and Communication System, also uses two-way range determination, see e.g. [2]. In this case, the control center sends a signal to potential users via a satellite transponder. A user, who wishes to be positioned answers with a predefined delay, via the same path. The setup is thus inverse to PRARE.

The main difficulty with range measurements is that they require two way transmissions. Such an approach is neither feasible in the civil area with billions of terminals nor desirable in the military domain, since any transmission would disclose the presence of the equipment. In both cases, receive-only equipments are needed. Thus, GNS Systems use a different approach.

3.1.2 Position from Pseudoranges

In GNS Systems the satellite encodes the transmit time into the signal, and the receiver determines the receive time of the signal. Both the satellite and the receiver use their own local clocks. The satellite clocks are carefully synchronized by the system's ground segment, while the receiver's clock-offset must be determined in the estimation process. This implies that four measurements are needed to determine the four unknown: x, y, z, and t. Consider a wavefront transmitted in $\vec{r}^{\,k}$ at time t^k. This wavefront will propagate on a conus $\mathcal{C}^k$ - called light cone - whose points $\vec{r}, t$ satisfy the equation:

$$\|\vec{r} - \vec{r}^{\,k}\|^2 - c^2(t - t^k)^2 = 0. \tag{3.2}$$

Note the subtle difference between Equations (3.2) and (3.1). The parameter τ^k in Equation (3.2) is a measurement result, and is thus known. The variable t in Equation (3.2), on the other side, describes the fourth space-time dimension, and is unknown. It needs to be determined jointly with the three spatial variables $\vec{r}$. This explains the difference in the geometry of the solution. The circles and spheres from the previous section are expanded into coni in 2, 3, and 4 dimensional space-time.

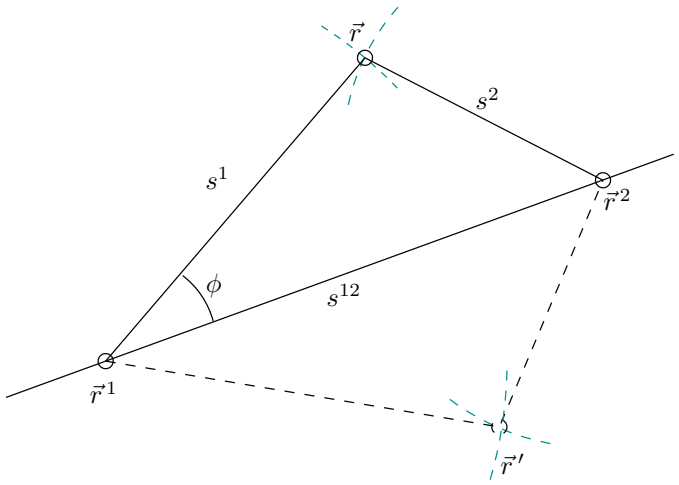

Figure 3.1: The position of $\vec{r}$ is determined by specifying the distances s^1, and s^2 from two reference points (satelites) in $\vec{r}^{\,1}$ and $\vec{r}^{\,2}$, respectively. There are two distinct solutions $\vec{r}$ and $\vec{r}\,'$ to this problem.

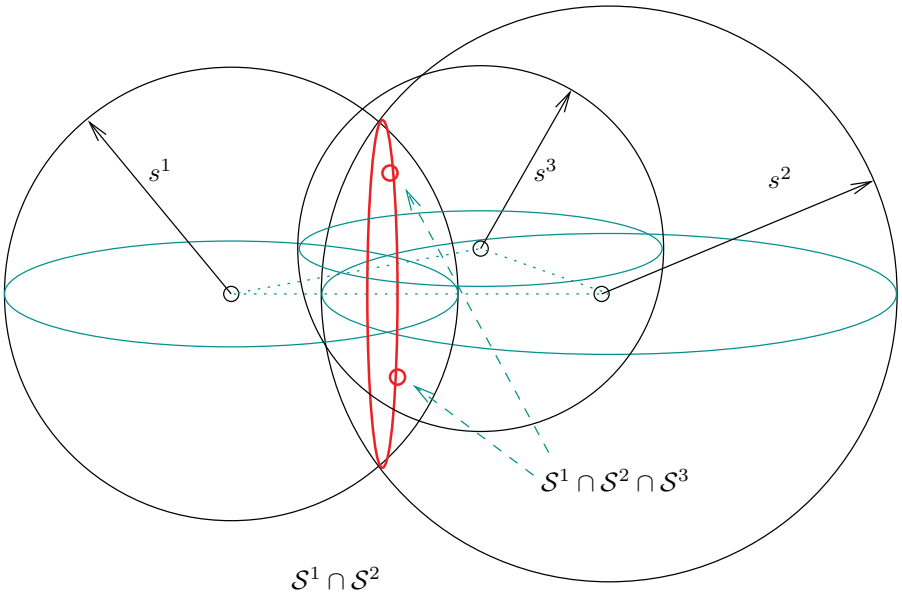

Figure 3.2: In three dimensions the position is determined by the intersection of three spheres $\mathcal{S}^1$, $\mathcal{S}^2$, and $\mathcal{S}^3$ of known radius s^k, centered around the reference point (satellites) in $\vec{r}^{\,k}$.

The simplest example is in one physical space dimension, i.e., in (1+1) dimensional space-time. The vector notation $\vec{r}$ is maintained, although trivial in one-dimension[2]. A space-time plot for the one-dimensional case is shown in Figure 3.3. The causal cones

$$\mathcal{C}^k_+ = \{(\vec{r},t) | t \geq t^k \text{ and } \|\vec{r} - \vec{r}^{\,k}\|^2 - c^2(t-t^k)^2 = 0\},$$

[2]The expressions are equally valid in higher dimensional situations.

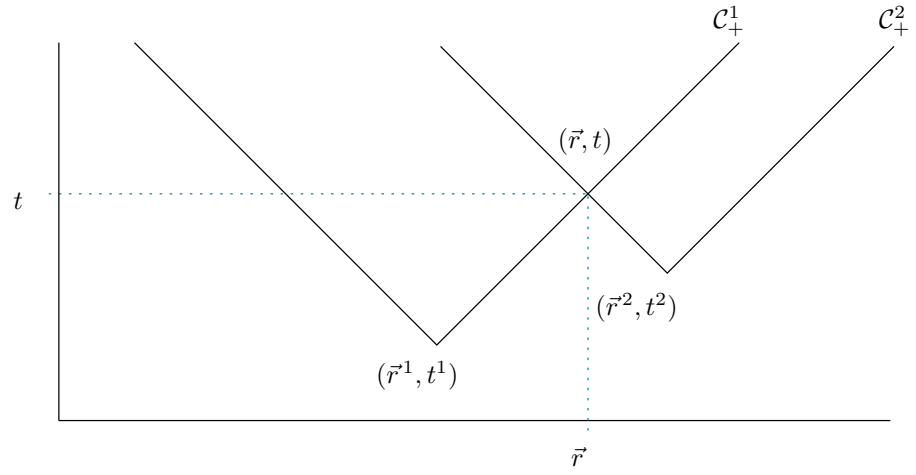

Figure 3.3: A wave transmitted from $x = 0$ at time $t = 0$ propagates along the axes $x = \pm ct$. It will intersect any other light cone, which has its origin in the acausal region of the first light cone in a unique point.

i.e., the part of the cone with $t \geq t^k$, which represents the propagation of the wave into the future is shown as a V. The length of the axes is scaled in such a manner that waves propagate on the diagonals[3]. The intersection of two cones in 2-dimensional space-time is an event with a space coordinate $\vec{r}$ and a time coordinate t. The space coordinate $\vec{r}$ is the desired user location. The time coordinate t is the time at which the two signals meet at the user location. This allows the receiver to adjust its clock to system time. Users who travel with different velocities will use different coordinate systems - see the later chapter on relativity. The light cones stay identical, at least in the absence of extreme gravitational fields. This invariance holds in the surrounding of the earth to a very high degree. A navigation system at the border of a black hole would need to consider the distortions of the metric caused by gravitational fields, however.

The interior of $\mathcal{C}_+^k$ is called the causal region $\mathcal{R}_+^k$. Events in

$$\mathcal{R}_+^k = \{(\vec{r}, t) | t \geq t^k \text{ and } \|\vec{r} - \vec{r}^k\|^2 - c^2(t - t^k)^2 < 0\}$$

can be influenced by events in $(\vec{r}^k, t^k)$. In a similar way one defines the acausal region

$$\mathcal{A}^k = \{(\vec{r}, t) | \, \|\vec{r} - \vec{r}^k\|^2 - c^2(t - t^k)^2 > 0\}$$

and for completeness

$$\mathcal{C}_-^k = \{(\vec{r}, t) | t \leq t^k \text{ and } \|\vec{r} - \vec{r}^k\|^2 - c^2(t - t^k)^2 = 0\},$$

$$\mathcal{R}_-^k = \{(\vec{r}, t) | t \leq t^k \text{ and } \|\vec{r} - \vec{r}^k\|^2 - c^2(t - t^k)^2 < 0\}.$$

These different regions are shown in Figure 3.4. Wavefronts from $\mathcal{R}_-^k$ and $\mathcal{R}_+^k$ can never reach $\mathcal{C}_+^k$, this would require that they travel faster or slower than the velocity of light, which is both not possible. Correspondingly, wavefronts transmitted in $(\vec{r}^k, t^k)$ and $\mathcal{R}_-^k \cup \mathcal{R}_+^k$ will never meet! A wavefront transmitted from $\mathcal{C}_+^k$, on the other side will either stay with the wavefront transmitted in $(\vec{r}^k, t^k)$ effectively reinforcing its energy if it travels in the same direction or never meet if it travels in a different direction. All these situations are not very useful for navigation. Therefore the only

[3]If the time axis has ticks at full seconds, the space dimension needs to have ticks at 299792458 meters.

useful location for other satellites is the acausal region $\mathcal{A}^i$. This requirement is a generalization of the condition $s^{12} \leq s^1 + s^2$ in the two dimensional ranging task.

Figure 3.3 makes it obvious that $(\vec{r}, t)$ can be determined from $(\vec{r}^1, t^1)$ and $(\vec{r}^2, t^2)$. The algebra is trivial. The solution at the intersection of the forward coni of the two satellites $\mathcal{C}^1_+ \cap \mathcal{C}^2_+$ is unique. This is in contrast to the ranging problem. A second solution $\mathcal{C}^1_- \cap \mathcal{C}^2_-$ exists but is acausal. It would apply if the waves were propagating backward in time.

The only missing element for computing the receiver's position are the space time coordinates $(\vec{r}^1, t^1)$ and $(\vec{r}^2, t^2)$ of the transmit events. They are provided by modulating the necessary information onto the ranging signals, which are continuous spread-spectrum signals. The latter information contains the satellite's orbit and the time offset of their clocks with respect to a global time reference - called system time.

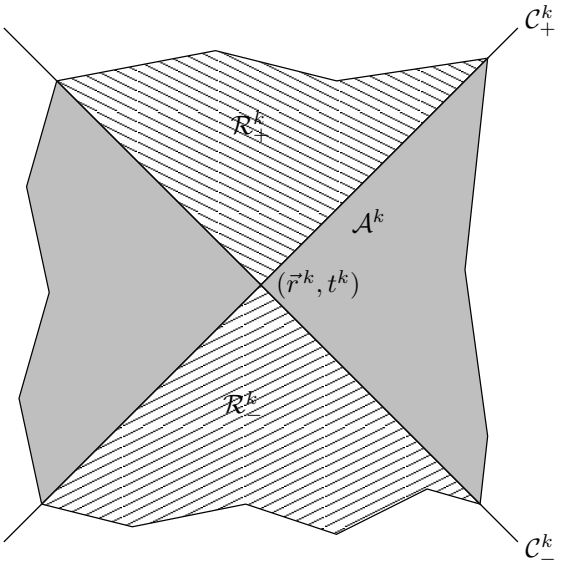

Figure 3.4: A forward and backward light cone $\mathcal{C}^k_+$ and $\mathcal{C}^k_-$ is attached to every space-time point $(\vec{r}^k, t^k)$. Events in $\mathcal{R}^k_+$, i.e. the interior of $\mathcal{C}^k_+$, can be influenced by $(\vec{r}^k, t^k)$, while $(\vec{r}^k, t^k)$ can be influenced by events from $\mathcal{R}^k_-$. The acausal region $\mathcal{A}^k$ is not related to $(\vec{r}^k, t^k)$.

3.1.3 The Case of Three Space-Time Dimensions

The picture in 1+1 space-time dimensions can be generalized to 2+1 space-time dimensions. In this case, the receiver's position is at the intersection of three coni $\mathcal{C}^1_+ \cap \mathcal{C}^2_+ \cap \mathcal{C}^3_+$ (see Figure 3.5). The geometry is most easily described by choosing a coordinate system such that the first satellite transmits at time 0, while being at the origin of the coordinate system. The second satellite transmits at time t' while being on the y-axis:

$$\vec{r}^1 = \vec{0},\ t^1 = 0; \quad \text{and} \quad \vec{r}^2 = (0, y'), t^2 = t'.$$

In Figure 3.5, the choice $y^3 \geq y^2 \geq y^1$ was furthermore made. The two coni associated with the above transmission events are described by:

$$\begin{aligned} \mathcal{C}^1_+ &: \quad x^2 + y^2 = (ct)^2, \quad t \geq 0 \\ \mathcal{C}^2_+ &: \quad x^2 + (y - y')^2 = c^2(t - t')^2. \end{aligned}$$

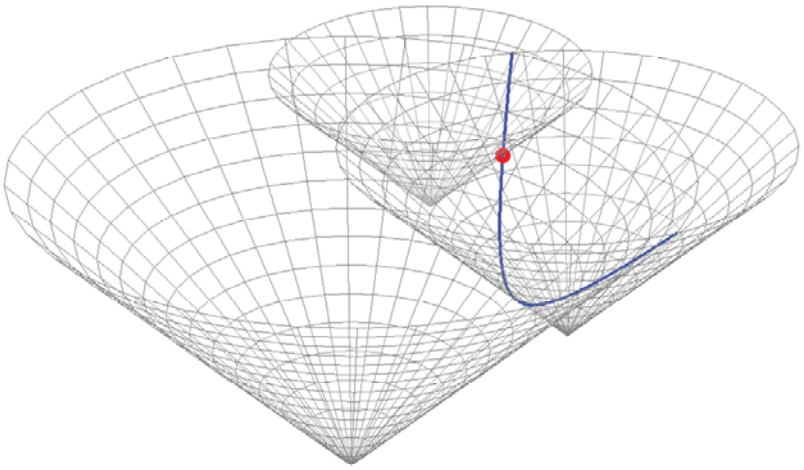

Figure 3.5: The space-time point describing the receiver time and position is at the intersection of three coni. The intersection of the two coni that start intersecting first is a hyperbola. The intersection of that hyperbola with the third conus leads to the desired point.

The difference of the two equations leads to the condition

$$y = \frac{y'}{2} + \frac{ct'}{y'}\left(ct - \frac{ct'}{2}\right) = \alpha + \beta ct.$$

This implies that:

$$\mathcal{C}_+^1 \cap \mathcal{C}_+^2 \subset \{(x, y, t) | y = \alpha + \beta ct\},$$

which is a plane parallel to the x-axis. $\mathcal{C}_+^1 \cap \mathcal{C}_+^2$ becomes the intersection of the cone $\mathcal{C}_+^1$ with that plane, i.e., a conical section. The slope and offset of the plane are given by

$$\beta = \frac{ct'}{y'}, \ \alpha = \frac{y'}{2}(1 - \beta^2).$$

The condition that the second satellite transmits its wavefront in the acausal region of the transmission by the first satellite implies that $y' > ct'$, and therefore, $\beta < 1$ and $\alpha > 0$. As a consequence, the intersection of the plane with the conus $\mathcal{C}_+^1$ (and also $\mathcal{C}_+^2$) is a hyperbola. This hyperbola is finally intersected with the last cone $\mathcal{C}_+^3$. The solution is unique if the third satellite is not on the y-axis, i.e., if the satellites are not co-linear.

This geometric representation becomes difficult to draw with three space dimensions. It is correspondingly useful to return to the picture of the previous section, and to consider a succession of instants, i.e. cuts of space-time in the space direction. Figure 3.6 shows three cuts of the coni, described in more details in the caption of the figure. The events of transmission of the satellite number 2 and 3 are made visible in the individual pictures, as well as the propagation of the wavefronts. Note that the position is unique whenever the satellites are not aligned, i.e., when

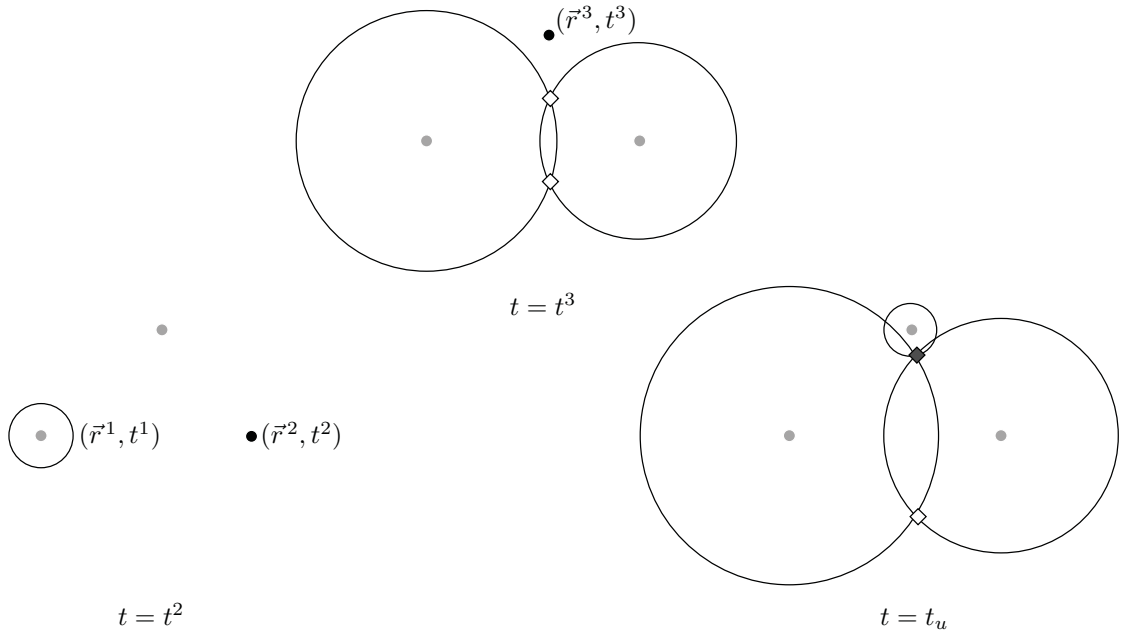

Figure 3.6: The first picture shows a snapshot at time t^2, the time at which the second satellite transmits, while the wave transmitted by the first satellite at time t^1 propagates. In the second picture, the wavefront of the pulses transmitted by Satellite 1 and 2 have started intersecting and Satellite 3 is transmitting. Finally, the last picture shows the instant at which all three wavefront intersect. This happens at the receiver's location.

they span 2D position space. The satellites are free to move. The receiver only needs to know the satellites' position at the time of transmission.

The latter picture can immediately be translated into 4D space time. Two spheres intersect in a circle, which expands as time evolves. The intersection with the third sphere becomes a pair of points, whose separation increases with time. The intersection with the fourth sphere determines the location of the receiver and the time of reception of the four signals. Again the solution is unique if the satellites are not located in the same plane.

Algebraic solutions of the non-linear equations for position and time have been derived by Grafarend and Shan [3]. In practical situations, the signal generation, the propagation and the measurements are all noisy. Furthermore, measurements from more than four satellites are available, which implies that the solution would in principle be overdetermined.

3.2 Solutions in the Presence of Noise

GNSS receivers compute the difference between their clock reading at the time of reception of the signal t' minus the time of transmission of that signal at the satellite t'^k. The primes indicate that time is measured using the respective clocks of the receiver and satellite k. If all clocks were synchronized, the difference would directly provide the range. Since this is not the case, a new name for these quantities was introduced: the difference $t' - t'^k$ is called "pseudorange". Remember that the transmit time t'^k of each satellite is encoded in its signal.

The reception process corrupts the signal by additive noise, since the antenna captures disturbances from the background and since the receiver operates at room temperature. In addition, the transmitter clock adds noise. The same holds for propagation, which includes e.g. scintillations

in the ionosphere, and diffuse multipath near the receiver. Some of these components add contributions to the signal amplitude others directly corrupt the pseudorange. The noise on the signal amplitude also causes noise in the delay estimation. The overall pseudorange noise from both types of contributions is modeled as additive Gaussian noise. The noise contributions on different pseudoranges η^k are not necessarily independent. Ionospheric and tropospheric contributions may affect the signal of several satellites, for example. The noise is assumed to be Gaussian with zero mean and a covariance given by:

$$C^{kl} = \mathcal{E}[\eta^k \eta^l]. \tag{3.3}$$

Assuming that the noise is zero-mean is justified since biases due to satellite clocks, ionospheric delay, tropospheric delay, multipath, and instrumental biases of the receiver are included in the deterministic corrections discussed in the chapters on time, propagation, and signals.

The mathematical definition of the noisy pseudorange is thus given by[4]

$$\tilde{\rho}^k = c(t' - t'^k) + \eta^k. \tag{3.4}$$

As already stated, it is the difference of the time of arrival of the signal t' measured using the receiver's clock and the time of the transmission at the satellite t'^k measured using the satellite's clock. The expression can be expanded, and simplified using Equation (3.2):

$$\begin{aligned} \tilde{\rho}^k &= c(t' - t + t - t^k + t^k - t'^k) + \eta^k \\ &= c(t' - t) + \|\vec{r} - \vec{r}^k\| - c(t'^k - t^k) + \eta^k. \end{aligned} \tag{3.5}$$

In this expression t is the time of arrival, and t^k is the time of transmission at satellite k, both expressed in system time. The difference $\delta^k = t'^k - t^k$ is determined by the ground segment and provided in the navigation message. The difference $\delta = t' - t$ is specific to the receiver but unknown to him, and must be determined jointly with the position. This is the fourth variable requiring the presence of a fourth satellite. With these notations, Equation (3.5) reads:

$$\tilde{\rho}^k = \|\vec{r} - \vec{r}^k\| + c(\delta - \delta^k) + \eta^k. \tag{3.6}$$

This equation is a non-linear stochastic equation. It is non-linear due to the presence of the norm $\|.\|$, and stochastic due to the presence of noise. Starting from first principle, would mean to search for the the most likely value $\xi^T = (\vec{r}^T, c\delta)$, given a certain set of pseudorange measurements $(\tilde{\rho}^1, \ldots, \tilde{\rho}^K)$. For Gaussian noise, this leads to a system of non-linear equations, which can be solved iteratively by the Newton-Raphson method, as described in Appendix B to Chapter 1. The Newton-Raphson method requires that the system of Equations (3.6) for $k \in \{1 \ldots K\}$ with K being the number of visible satellites be linearized. This can be done in a straight forward manner (see Exercise 3). Alternatively, the non-linear norm can be reinterpreted as follows:

$$\begin{aligned} \|\vec{r} - \vec{r}^k\| &= \frac{(\vec{r} - \vec{r}^k)^T}{\|\vec{r} - \vec{r}^k\|}(\vec{r} - \vec{r}^k) \\ &= (\vec{e}^k)^T(\vec{r} - \vec{r}^k), \end{aligned}$$

with $\vec{e}^k$ being the unit vector pointing from the satellite towards the receiver. Since the satellites fly on orbits that have a radius of 26600-29600 km, while the receivers are near the surface of the earth at a radius of 6378 km, the vector $\vec{e}^k$ is not very sensitive to the receiver's location. Typically the Newton-Raphson iteration is started assuming that the receiver is at the center of the earth. The iterations then converges to the desired level of accuracy in 4 -5 steps! If n is the index of the iteration, the linearized equations in the n-th step are:

$$\tilde{\rho}^k = (\vec{e}^k_{n-1})^T(\vec{r}_n - \vec{r}^k) + c(\delta_n - \delta^k) + \eta^k \tag{3.7}$$

[4]The variable $\tilde{\rho}$ denotes the pseudorange, while ρ will be reserved for the pseudorange corrected by the satellite clock offset and satellite location. The latter is a convention that is widely used.

with $\vec{r}_n$ and δ_n being the position and clock offset of the receiver obtained in the n-th iteration, and

$$\vec{e}^{\,k}_{n-1} = \begin{cases} -\dfrac{\vec{r}^{\,k}}{\|\vec{r}^{\,k}\|} & \text{if } n = 1, \\ \dfrac{(\vec{r}_{n-1} - \vec{r}^{\,k})}{\|\vec{r}_{n-1} - \vec{r}^{\,k}\|} & \text{if } n > 1. \end{cases}$$

Equation (3.7) is linear and can be solved for $\xi_n^T = (\vec{r}_n^{\,T}, c\delta_n)$, under conditions developed later in this section. The index n is not written out any further in order to simplify the writing. It is present implicitly and could be restored at any time. The introduction of matrix notation serves the same purpose. Note that there are different types of "vectors:" the k-tuples of measurements ρ and noise η, and the four-dimensional space-time vectors ξ. With these notations, Equation (3.7) becomes:

$$\rho = H\xi + \eta \tag{3.8}$$

with

$$\rho = \begin{pmatrix} \tilde{\rho}^1 + (\vec{e}^{\,1})^T\vec{r}^{\,1} + c\delta^1 \\ \vdots \\ \tilde{\rho}^K + (\vec{e}^{\,k})^T\vec{r}^{\,k} + c\delta^K \end{pmatrix}, \quad H = \begin{pmatrix} (\vec{e}^{\,1})^T & 1 \\ \vdots & \vdots \\ (\vec{e}^{\,K})^T & 1 \end{pmatrix}, \quad \xi = \begin{pmatrix} \vec{r} \\ c\delta \end{pmatrix}, \quad \eta = \begin{pmatrix} \eta^1 \\ \vdots \\ \eta^K \end{pmatrix}, \tag{3.9}$$

and a noise covariance given by Equation (3.3). In this expression, the vector ρ consists of corrected components, with clock, and orbit corrections $c\delta^k$, and $(\vec{e}^{\,k})^T\vec{r}^{\,k}$, obtained from the navigation message. This again simplifies the notations but has to be kept in mind.

Under the assumption of the linear model of Equation (3.8), and the noise model of Equation (3.3), the probability of a measurement ρ, given that the receiver has space-time coordinates ξ, becomes:

$$\begin{aligned} p(\rho|\xi) &= p(\eta) \\ &= \frac{1}{(2\pi)^{K/2}\sqrt{\det(C)}}\, e^{-\frac{1}{2}\eta^T C^{-1}\eta} \\ &= \frac{1}{(2\pi)^{K/2}\sqrt{\det(C)}}\, e^{-\frac{1}{2}(\rho - H\xi)^T C^{-1}(\rho - H\xi)}. \end{aligned}$$

The result can be used in a maximum likelihood or a maximum a posteriori estimate. In the simpler maximum likelihood estimation, the receiver chooses the value of ξ that has led to the measured ρ with the highest probability. The estimate then becomes:

$$\hat{\xi} = \arg\max_{\xi} p(\rho|\xi).$$

The mapping of $p(\rho|\xi)$ by a monotonic function will not change the result, and another mapping by changing the sign of the expression will also lead to the same result provided that "taking the maximum" is replaced by "taking the minimum." Thus the negative log-likelihood must be minimized:

$$\begin{aligned} \hat{\xi} &= \arg\min_{\xi} \left(-\log p(\rho|\xi)\right) \\ &= \arg\min_{\xi} (\rho - H\xi)^T C^{-1} (\rho - H\xi), \end{aligned}$$

which is a weighted least-squares problem. This can be written in a slightly more compact manner by defining:

$$\|\zeta\|^2_{C^{-1}} = \zeta^T C^{-1} \zeta, \tag{3.10}$$

and further simplified by introducing the weighting matrix $W = C^{-1}$. The maximum likelihood estimate then becomes:

$$\hat{\xi} = \arg\min_{\xi} \|\rho - H\xi\|^2_W. \tag{3.11}$$

For the maximum a posteriori estimation, we need to apply Bayes' rule for determining the joint probability of ρ and ξ :

$$p(\rho, \xi) = p(\rho|\xi)p(\xi),$$

with $p(\xi)$ being the a priori knowledge about ξ. Such prior information might be available from previous measurements, which is typically handled in a Kalman filter or due to external constraints. In this case, the maximum a posteriori probability becomes

$$p(\xi|\rho) = \frac{p(\rho|\xi)p(\xi)}{p(\rho)}, \tag{3.12}$$

with a normalization given by

$$p(\rho) = \int_{\mathbb{R}^4} d^4\xi' p(\rho|\xi')p(\xi').$$

Soft external constraints were introduced in relative positioning recently by Jurkowski, Henkel, Gao, and Günther [4] to characterize receivers at the two ends of a hanging rope. Assuming a Gaussian prior of the form:

$$p(\xi) = \frac{1}{\sqrt{(2\pi)^4 \det D}}\, e^{-1\frac{1}{2}(\xi-\xi_0)D^{-1}(\xi-\xi_0)},$$

with ξ_0 being the expected value, and with D being the covariance of the a priori information on ξ, implies that the maximization of Equation (3.12) becomes equivalent to the minimization of

$$\|\rho - H\xi\|^2_{C^{-1}} + \|\xi - \xi_0\|^2_{D^{-1}}. \tag{3.13}$$

In the important case of no a priori knowledge, the maximization of the log-likelihood and of the a posteriori probability are equivalent.

The solution of the log-likelihood problem or equivalently of the least-squares problem formulated in Equation (3.8) is derived in most textbooks on signal processing and estimation theory, like Kay [5] for example. For completeness, a derivation of this central result is provided in Appendix A as well. The covariance matrix C is non-singular, since none of the measurements is noiseless, and since they are rather uncorrelated. In this case $W = C^{-1}$ is also non-singular.

The solution to Equation (3.11), i.e. the estimate $\hat{\xi}$ of ξ is then given by (see Appendix A):

$$\hat{\xi} = (H^T W H)^{-1} H^T W \rho, \tag{3.14}$$

provided that $H^T W H$ is non-singular. This means that the least-squares problem has a unique solution that can be computed using linear algebra[5]. In the context of the Newton-Raphson method, this solution might be written in the form

$$\hat{\xi}_n = (H^T_{n-1} W H_{n-1})^{-1} H^T_{n-1} W \rho_n,$$

with the estimate $\hat{\xi}_{n-1}$ being used in the computation of H_{n-1}, and of the K-tuple ρ_n. The latter tuple has the components $\tilde{\rho}^k + (\vec{e}^k_{n-1})^T \vec{r}^k + c\delta^k$. After a small number of iteration $\hat{\xi}_n$ converges to a limiting value $\hat{\xi}$. This limiting value is a solution of Equation (3.14).

A rectangular $K \times 4$ matrix H is defined to be non-singular iff:

$$H\xi = 0 \quad \text{implies} \quad \xi = 0.$$

Every non-singular square matrix G has an inverse G^{-1} with the property that $G^{-1}G = GG^{-1} = 1$. In the present case, $G = H^T W H$. It can easily be verified that "H singular" is equivalent to "G

[5]For notational convenience, the hat is not written when it is clear that the estimate is considered - objects like $\hat{\vec{r}}$ look clumsy.

singular." The direction "H singular $\Rightarrow$ G singular" is trivial. For the other direction, assume that G is singular, then there is $\xi \neq 0$ with $H^T W H \xi = 0$, and therefore $\xi^T H^T W H \xi = \|H\xi\|_W^2 = 0$. Since W is non-singular, this implies $H\xi = 0$, i.e. H is singular.

The condition of H being (non-)singular can be interpreted in geometrical terms. If H is singular, this implies that there is a vector $\xi^T = (\vec{r}^T, c\delta) \neq 0$ such that

$$(\vec{e}^k)^T \vec{r} + c\delta = 0, \ \forall\ 1 \leq k \leq K,$$

i.e. that there is a vector $\vec{r}$ such that all unit vectors $\vec{e}^k$ pointing from the satellites to the receiver form the same angle with $\vec{r}$, i.e. their beginning lies in a plane, and since they are unit vectors, also on a circle (see Figure 3.7). With $K = 3$ satellites such an $\vec{r}$ always exists, since the ends of

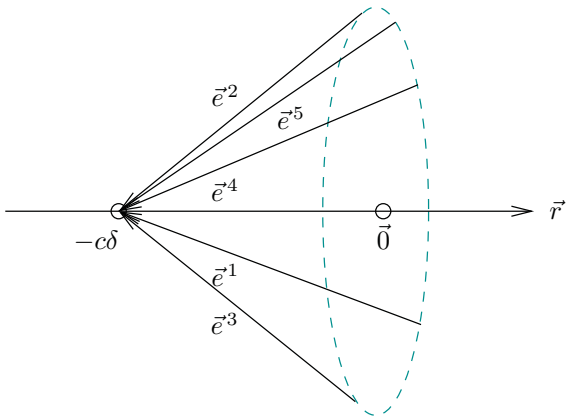

Figure 3.7: The square matrix $H^T W H$ is singular iff there is a $\vec{r}$ such that all unit vectors $\vec{e}^k$ lie on a circle around $\vec{r}$.

$\vec{e}^1, \vec{e}^2, \vec{e}^3$ define a plane. The vector $\vec{r}$ is then normal to that plane. For $K = 4$, it is very unlikely to occur, and with an increasing number of satellites K it becomes even less likely (a generic chair with four legs and more rocks). In a real situation, the interesting question is not only to know whether H is singular, but also how close it is to being singular. An appropriate concept is the "geometric dilution of precision" discussed next.

3.3 Geometric Dilution of Precision

The impact of measurement noise on position and time estimation are described next. A simple two-dimensional example from ranging, shown in Figure 3.8, is suitable for developing intuition. The curves in that figure describe lines of equal distance from the satellites. The uncertainty in the ranges implies that the actual distance lies in a region enclosed by two such curves, e.g. with a probability of 0.63 (1σ) in the case of Gaussian noise. The receiver is most likely at the intersection of two such regions. The size of this intersection is proportional to $1/\sin\alpha$, which grows without bound when $\alpha \to 0$. This is valid for both the maximum linear dimension $\epsilon/\sin\alpha$, and the surface $\epsilon^2/\sin\alpha$. The angle $\alpha = 0$ corresponds to collocated satellites, which is a singular situation.

The estimate of position and clock offset in satellite navigation is the linear function of the pseudoranges described by Equation (3.14). The additive white Gaussian nature of the noise of the pseudorange measurement ρ implies that the noise of the position and clock offset $\hat{\xi}$ is also

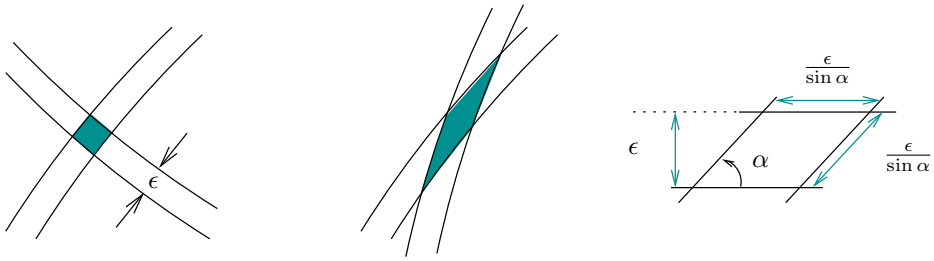

Figure 3.8: Both drawings show measurements that are affected by noise of a similar variance. The geometry on the right side leads to a significantly larger position uncertainty.

additive white Gaussian:

$$\begin{aligned}\hat{\xi} &= (H^T W H)^{-1} H^T W \rho \\ &= \xi + \eta_\xi,\end{aligned}$$

with

$$\eta_\xi = (H^T W H)^{-1} H^T W \eta.$$

Since the mean of η was assumed to be zero, this implies

$$\mathcal{E}[\eta_\xi] = 0.$$

Similarly

$$C_\xi = \mathcal{E}[\eta_\xi \eta_\xi^T] = (H^T W H)^{-1} = (H^T C^{-1} H)^{-1}, \tag{3.15}$$

which provides an expression for the covariance C_ξ of the position and clock offset ξ in terms of the covariance C of the measurements ρ. This expression is equal to the factor occurring in the least-squares solution (Equation (3.14)), which is undetermined whenever the matrix $H^T W H$ is singular. Thus the scenario for running into an undetermined solution, is that some component of the covariance matrix C_ξ becomes infinite. The design of actual constellations is such that this does not occur, at least not in an open sky environment.

In the case $C = \sigma^2 \mathbf{1}$, the above equation has the interesting property to separate the range and clock error into a purely geometric part $(H^T H)^{-1}$ - denoted as Dilution of Precision (DOP) - and a part that only depends on the pseudorange uncertainty σ - called User Equivalent Range Error (UERE). Define

$$H = (h_x, h_y, h_z, h_1),$$

with

$$h_a = (e_a^1 \dots e_a^K)^T \quad \text{for } a \in \{x, y, z\}, \quad \text{and} \quad h_1 = (1 \dots 1)^T,$$

then

$$H^T H = \begin{pmatrix} h_x^T h_x & h_x^T h_y & h_x^T h_z & h_x^T h_1 \\ h_y^T h_x & h_y^T h_y & h_y^T h_z & h_y^T h_1 \\ h_z^T h_x & h_z^T h_y & h_z^T h_z & h_z^T h_1 \\ h_1^T h_x & h_1^T h_y & h_1^T h_z & h_1^T h_1 \end{pmatrix},$$

and

$$h_a^T h_b = \sum_{k=1}^{K} h_a^k h_b^k, \quad h_a^T h_1 = \sum_{k=1}^{K} h_a^k, \quad h_1 h_1 = K,$$

for $a, b \in \{x, y, z\}$. The matrix $H^T H$ is thus associated with metrics, in which indices run over satellites. The matrix inversion in $(H^T H)^{-1}$ mixes these components in a manner that typically requires numerical evaluation.

Users are primarily interested in the precision of their horizontal, and vertical position, as well as of their time offset. For this reason the DOP is typically described in a local tangential coordinate system, with axis pointing into the directions East, North and Up (see Figure 3.9), as well as in the time direction. This leads to the definition of the following dilutions of precision: EDOP (East),

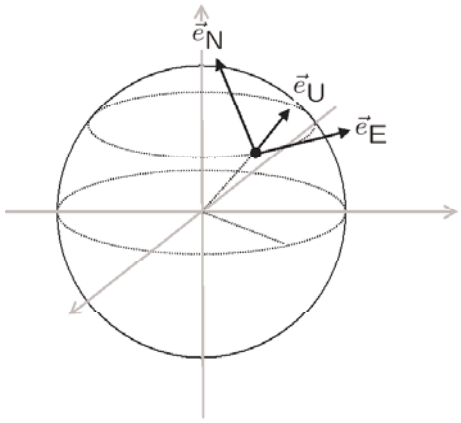

Figure 3.9: Both drawings show measurements that are affected by noise of a similar variance. The geometry on the right side leads to significantly larger position uncertainty.

NDOP (North), VDOP (Vertical/Up), and TDOP (Time). They are obtained by transforming the matrix $(H^T H)^{-1}$ into the tangential ENU coordinate system through an appropriate rotation R[6]:

$$R(H^T H)^{-1} R^T =: \begin{pmatrix} \mathrm{EDOP}^2 & \cdot & \cdot & \cdot \\ \cdot & \mathrm{NDOP}^2 & \cdot & \cdot \\ \cdot & \cdot & \mathrm{VDOP}^2 & \cdot \\ \cdot & \cdot & \cdot & \mathrm{TDOP}^2 \end{pmatrix}.$$

Let $H_{\vec{e}}$ be the H-matrix obtained by using the vectors $\vec{e}^{\,k}$, then a little algebra leads to the identity

$$R(H_{\vec{e}}^T H_{\vec{e}})^{-1} R^T = (H_{R\vec{e}}^T H_{R\vec{e}})^{-1},$$

which shows that the DOP is based on the same type of metrics as described above. Since the $\vec{e}$-vectors (depend on the satellites' and user's movements) as well as the coordinate transformation R (only depends on the user's movement) are slowly varying, the same applies to DOP-values.

Important error measures are: the horizontal error σ_H, the vertical error σ_V, the position error σ_P, the time error σ_T, and the total error σ_G. They are characterized by

$$\sigma_X = \sigma\,\mathrm{XDOP}, \tag{3.16}$$

with

$$\mathrm{HDOP} = \sqrt{\mathrm{EDOP}^2 + \mathrm{NDOP}^2},$$
$$\mathrm{PDOP} = \sqrt{\mathrm{HDOP}^2 + \mathrm{VDOP}^2}, \quad \text{and} \quad \mathrm{GDOP} = \sqrt{\mathrm{PDOP}^2 + \mathrm{TDOP}^2}.$$

Equation (3.16) shows that the standard deviation is inflated by the dilution of precision, and thus that large values of DOPs have a negative impact precision.

[6]The three dimensional rotation $R_{3\times3}$ is expanded to four dimensions by keeping that dimension invariant, i.e. the matrix is bordered by zeros in the off-diagonal positions of the fourth dimension, and by one in the diagonal position.

The dilution of precision is a function of the relative geometry of the user and the satellites locations. It depends on the user position, the constellation geometry, the number of satellites, and the masking angle. The masking angle is the elevation angle above which satellites are used for computing the position of the receiver. In geodetic environments, typical values are 5-10 degrees. The signals of satellites at lower angles are discarded, due to the increasing degradation by multipath. A rule of thumb says that urban environments are typically associated with a masking angle of 30 degrees.

On ground, the values of the VDOP exceed those of the HDOP. Geometries that lead to good values of the HDOP are obtained with satellites, that are well distributed in azimuth, and are at low and moderate elevations. Since the earth is obstructing the path for satellites that are below the receiver, no equivalent geometries are available for achieving small values of the VDOP. In this case, the best geometries are with satellites at large and low elevations. Small values of the VDOP play a crucial role in aviation. The maximum values for the current GPS constellation, and the future Galileo constellation are plotted in Figure 3.10. The 27 satellites of Galileo are arranged in an optimum manner, and show a very consistent performance. The 30 satellites of the current GPS constellation (June 27th 2010) are arranged in a 24 satellite constellation with new satellites doubling orbital slots of older ones. The VDOP distribution for GPS shows the periodicity of half a sidereal day, i.e. the same patterns occur at two different longitudes.

Since the satellites are moving on their orbits, the dilution of precision is a function of time. Figure 3.11 shows the evolution of the H,V, and PDOP for 12 hours in Toulouse for GPS. The values vary widely, as satellites raise and set. The same applies to the TDOP (not shown). The plots also show that $H^T WT$ never becomes singular, and thus that a solution to Equation (3.14)always exists.

The split of the covariance C_ξ into a stochastic and a geometric factor has proven to be very useful in the analysis of constellation alternatives, in the planning of measurement campaigns and in the analysis of performance bounds as a function of location. Its limitations are mainly due to the dependence of multipath noise on elevation. Since satellites at low elevation play an important role in achieving low values of the DOP, this must be kept in mind.

3.4 Pseudorange Noise and Errors

Pseudorange noise and errors are caused by a number of phenomena. The importance of the individual components depends on the environment. Examples of environments range from the unobstructed sky seen by a flying aircraft to the street canyons of downtown Manhatten. The impact of the individual noise components also depends on the type of antenna and on the receiver used, in particular on the bandwidth, and the front-end design. The true noise processes are the receiver noise, the background thermal noise captured by the antenna, and the clock noise. Furthermore, diffuse multipath and ionospheric scintillation can also be treated as such. Ionospheric scintillations mainly occurs during ionospheric storms in periods of high solar activity. Orbital errors, ionospheric and tropospheric delays, as well as specular multipath may be measured or modeled. They are biases. Nevertheless their unmodelled contributions are typically also treated as noise, and more precisely as additive Gaussian noise. They are not white, however, and are typically not reduced in an averaging process.

In the following, the main contributions to noise and errors are shortly commented:

- *Ephemeris:* The ephemeris error is the error in the satellite orbit. Under nominal conditions, this error is dominated by the error in the determination, prediction and representation of the orbit in the control segment. The most accurate orbit determination is performed by the International GPS Service (IGS). Their post-processed orbits have an accuracy around 2 cm. The post-processing includes measurement data of orbital segments surrounding the

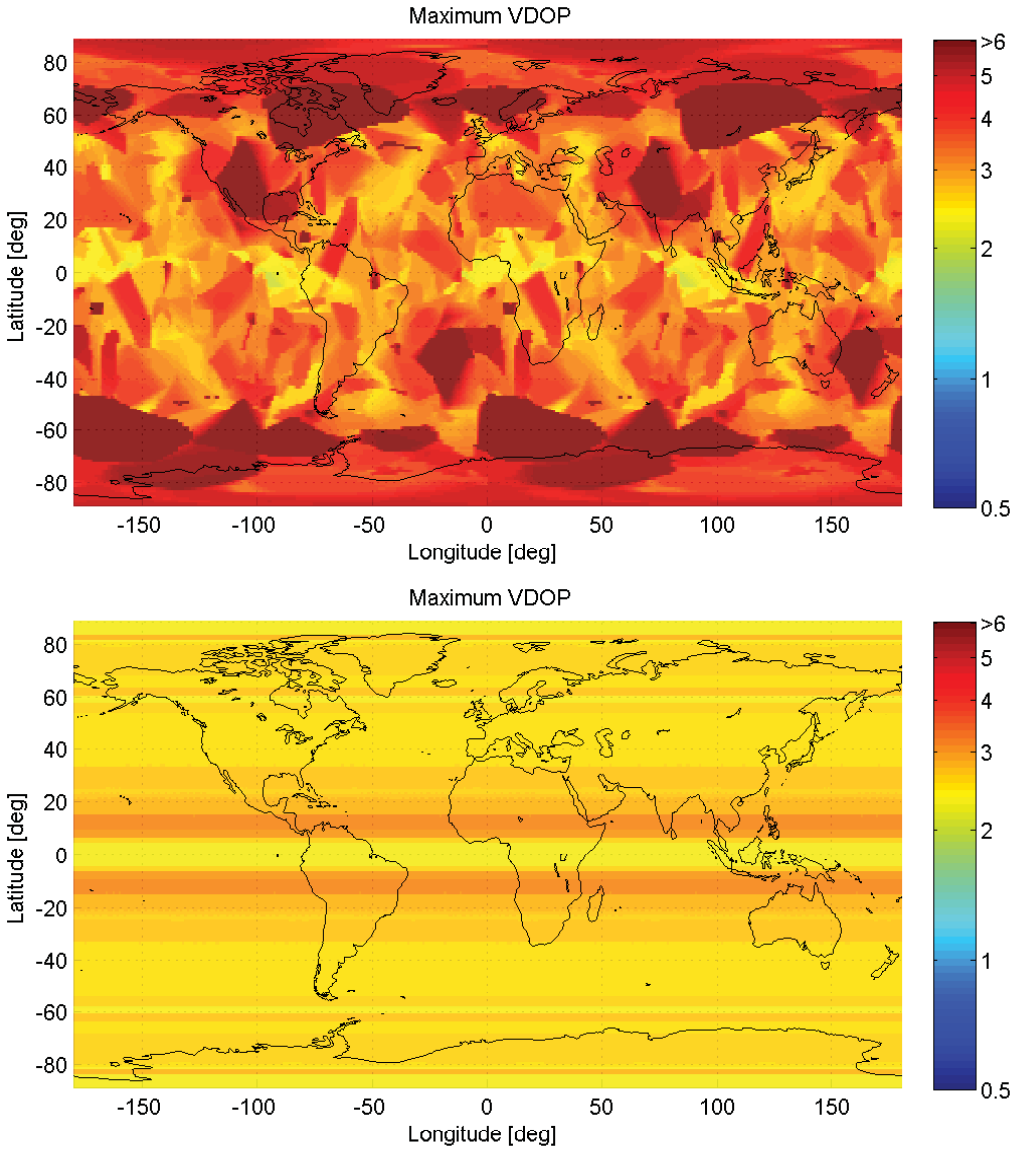

Figure 3.10: Maximum VDOP for the 30 satellite GPS in September 2010 and a 27 satellite Galileo constellation. [Courtesy: K. Giger, TUM 2010]

satellite position considered. IGS predicted orbits are reaching similar levels. The accuracy of the orbits broadcast by GPS are quite different for the component directed towards the center of the earth (radial), the component in the direction of the velocity vector (along track), and in the transversal direction (cross track). A study by Belabbas, [6], shows worst case radial components around 20 centimeters, along track errors up to 150 centimeters, and across track components in-between. The radial component has the largest impact on the pseudoranges. It is thus the component of highest interest to satellite navigation. Figure 3.12 plots the pseudorange error in Oberpfaffenhofen for a three-week time-span in 2009. This plot suggests that the ephemeris error models typically used for GPS with a σ of 1 meter is conservative. The value specified for Galileo is 45 centimeters. Abnormal errors can

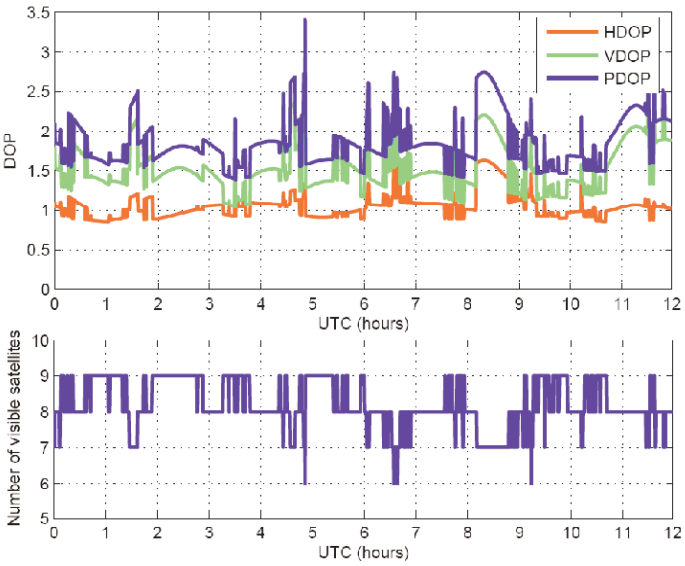

Figure 3.11: DOP values computed for Toulouse for the GPS constellation on October 4th, 2005. This pattern repeats every day as long as the constellation is not changed. Periods with fewer satellites lead to higher DOP values. The VDOP is roughly 1.5 times larger than the HDOP. [Courtesy: S. Graf, TUM 2006]

potentially occur if a thruster valve leaks or if a satellite is maneuvered without warning. The latter occurred once on April, 10th. 2007 [9].

- *Clock:* The inaccuracy of the satellite clocks has two causes: the uncertainty of the satellite position and the stochastic nature of the clocks. The former was discussed above, it directly impacts the accuracy of the determination of the satellite clock offsets by the ground segment. The clock parameters - bias, drift, and drift-rate - are determined by the control center and then broadcast during a period of validity of 120 minutes for GPS and 100 minutes for Galileo. They change over time in a stochastic manner. Modern Rubidium clocks lead to an uncertainty of around 0.3 ns in 100 Minutes (see Chapter 9), which corresponds to 10 cm. This allows to fulfill the specifications with margin. Clocks additionally show a number of abnormal behaviors, like time and frequency jumps or drifts. Since 2006, all GPS satellites are continuously monitored. One of the most significant error event occurred on January 1st, 2004, when the clock of satellite PRN023 run off. At the time of discovery, the pseudorange error had grown to 285 km [11].

- *Ionospheric Error:* Single frequency receivers correct the ionosphere by models with parameters provided by the system. The models are the Klobuchar-model for GPS and the NeQuick-model for Galileo. The residual zenith-error[7] of the Klobuchar model for a quiet ionosphere at mid latitude is shown in Figure 3.13. Typical standard variations are in the order of 5 meters. The ionospheric delay also depends on the season and on the phasing with respect to the 11 year solar cycle. In periods of high solar activity, 10 meters are a more realistic standard deviation. Under storm conditions, as well as in low and high latitudes, the delay can widely depart from the model and also change suddenly. This is the biggest

[7]The ionospheric and tropospheric errors can be separated into a geometric part, which describes the thickness of the atmospheric layer, and a zenith-delay, which characterizes the delay for the geometry with the lowest thickness.

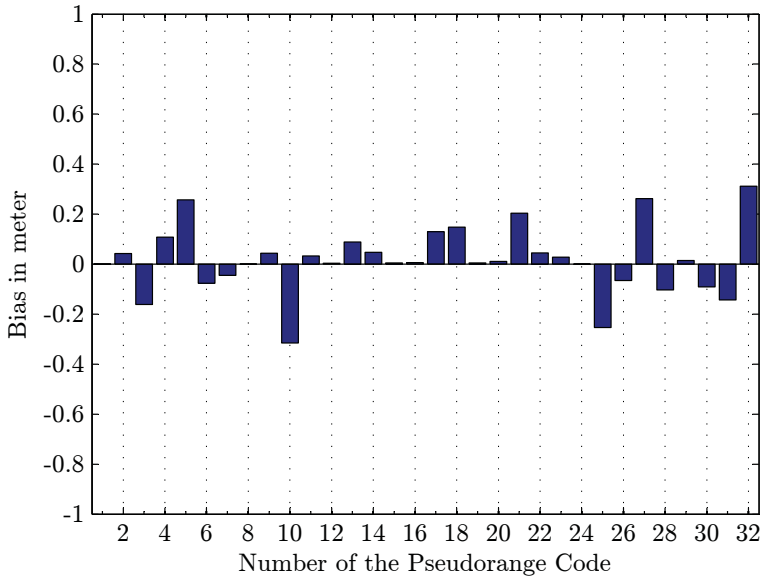

Figure 3.12: Pseudorange bias for Oberpfaffenhofen, obtained by comparing the broadcast ephemeris with the IGS orbits. The phase-center to center-of-gravity offsets were taken from the National Geospatial Intelligence Agency. The period of comparison was from October 11, 2009 00:00, until October 31, 2009 23:45. [Courtesy: Belabbas, DLR 2009].

threat to single frequency aviation-receivers, even in the case of Ground Based Augmentation Systems (GBAS) and Satellite Based Augmentation Systems (SBAS), see Chapter 13 and 14, respectively. Multifrequency open services are currently being introduced with GPS IIF/III, GLONASS, Galileo, and Beidou. These services mostly eliminate the ionosphere. Uncompensated higher orders effects are small.

- *Tropospheric Error:* The tropospheric delay can be surprisingly well described by models that only rely on the geographic latitude, the time of the day, and the time of the year. The use of measurement-dependent models, which are fed by ground measurements of the humidity, pressure, and temperature do not improve the performance substantially (see Chapter 7). Figure 3.14 shows the performance of the MOPS model for aviation [12] in Oberpfaffenhofen. The average performance of the model is $\sigma = 12$ cm. The measurement in Oberpfaffenhofen show a slightly worse performance.

- *Multipath:* Multipath propagation causes an offset in the pseudorange estimation, which can be very significant. In cities, reflections with excess delays of 50 meter and more often occur. The actual values are determined by the relative geometry of the satellite, the scatterers and the receiver. Sometimes components with large delays also occur in line-of-sight situations. An example of multipath that grows up to 100 meters is shown in Figure 3.15. The changing phase between the direct and multipath component are very helpful in cases in which the receiver is moving. In this case, the multipath error cancels out over time, at least to a certain extent. A technique, called carrier smoothed code, uses this property. It is described in Chapter 11.

- *Background and Receiver Noise:* Typical values of receiver noise are in the order of 0.8 meters

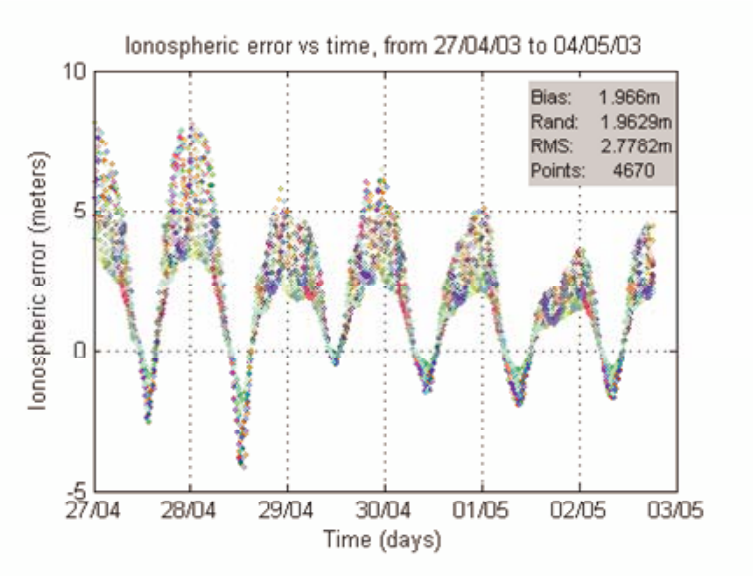

Figure 3.13: The residual zenith ionospheric error shows a strong daily variability. The plot shows a few days of data measured in Potsdam during Spring 2003. It shows that the ionosphere is quiet and particularly well modeled during night time [Courtesy: Belabbas, Petitprez, Hornbostel [7]].

for GPS today, and 0.3 and 0.08 meters for the Galileo L1 and E5a/b signals, as well as for GPS in the future (see Table 5.7 for details). Receivers are also subject to biases. The common bias on all signal components is included in the clock correction $c\delta$. Biases between frequencies, as well as between code and carrier must be estimated. First steps for estimating all relevant bias components have recently been published by Henkel, Wen, and Günther [10]. Receiver impairments can typically be detected by using several receivers in parallel.

In summary, nominal error components are typically modeled as zero-mean additive Gaussian noise of a given standard deviation. The superposition of several such contributions is again zero-mean additive Gaussian with a variance that is equal to the sum of the variances. This leads to the establishment of error budgets, like the one due to Kovach [8], shown in Table 3.1.

3.5 Velocity and Drift

In many applications, the velocity plays an important role. In aeronautical and maritime environments, for example, satellite navigation does not only provide the most accurate but also the simplest means for determining the velocity over ground. The velocity is defined by

$$\vec{v} = \frac{d}{dt}\vec{r} = \dot{\vec{r}}.$$

It is correspondingly natural to consider the formal time derivative of the pseudorange. Using the form of the pseudorange described by Equation (3.5), this becomes:

$$\dot{\rho}^k = \frac{d}{dt}\|\vec{r} - \vec{r}^{\,k}\| + c(\dot{\delta} - \dot{\delta}^k) + \dot{\vartheta}^k. \tag{3.17}$$

The derivative of the equation is formal for two reasons. Firstly, the pseudorange in Equation (3.7) is not an instantaneous pseudorange but an average over an interval of 1 ms up to one second in

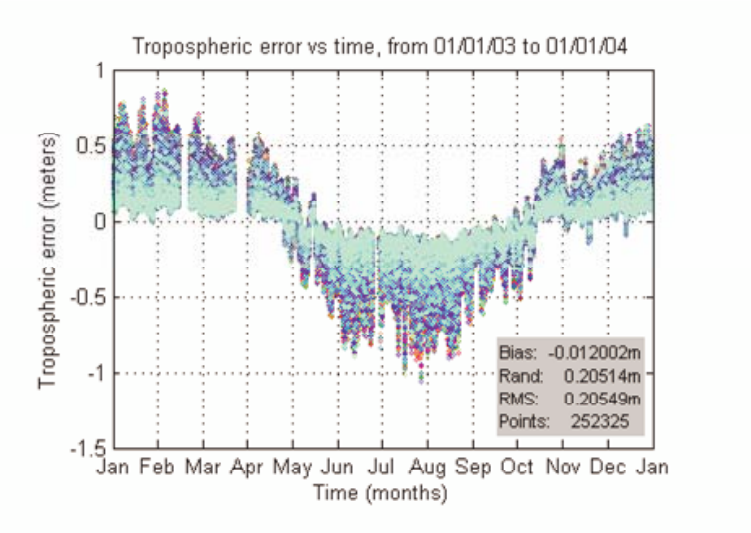

Figure 3.14: The residual zenith tropospheric error between the MOPS model of the FAA and the actual data determined by the IGS (SINEX) shows that the fitting for Oberpfaffenhofen could be slightly improved by reducing the delay in the winter and increasing it during summer [Courtesy: Belabbas, Petitprez, Hornbostel [7]].

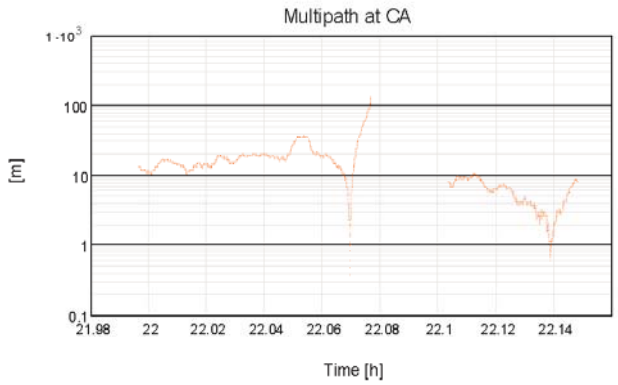

Figure 3.15: Extreme multipath in a line of sight situation in the Research Harbor of Rostock [Courtesy: E. Engler, DLR 2010].

extreme cases. The derivative in Equation (3.17) is defined as a difference of such measurements divided by the time separation of the measurements. Secondly, η^k in Equation (3.7) describes a stochastic process, whose derivative does typically not exist. The processes ϑ^k and η^k include different kinds of unmodeled errors relating to the orbits, clocks, propagation and receivers. An orbit error, for example, plays an important role in the position determination but is of marginal interest in velocity determination. The same applies to ionospheric and tropospheric errors. White frequency noise in the satellite clock - on the other hand - corrupts the velocity determination, while being averaged out in positioning. Thus ϑ^k cannot be determined from η^k but must be derived from first principles.

A most important aspect in velocity determination is that it can be based on carrier-phase measurements. Carrier phase measurements are much less sensitive to noise, as shall be developed

Allocation [m]	Single-Frequency C/A-Code Receiver	Dual-Frequency C/A-Code Receiver
Space Segment	*1.7*	*3.5*
Clock Stability	0.5	0.5
Group Delay Stability	1.6	1.1
Diff. Gr. Delay Stability	0	3.3
Other Satellite Errors	0.3	0.3
Control Segment	*5.5-10.1*	*3.7*
Clock/Eph. Estimation	0.8	0.8
Clock/Eph. Prediction	0.5	0.5
Clock/Eph. Representation	0.4	0.4
Clock/Eph. other	0.3	0.3
Ionospheric Delay	4.9-9.8	0
Group Delay Estimate	2.3	0
User Segment	*1.1*	*1.8*
Receiver noise	0.2	0.2
Multipath	0.9	0.9
Ionospheric Delay	0	1.4
Tropospheric Delay	0.5	0.5
Other UE Errors	0.4	0.4
Total System UERE	5.9-10.3	4.1

Table 3.1: Budget for the User Equivalent Range Error (UERE) as proposed by Kovach for a terrestrial user [8]. This budget is only indicative; it applies to an airplane on a runway. Multipath in a city would be significantly worse, for example. The UERE is dominated by the ionospheric delay in the case of single-frequency receivers and the uncertainty about the separation of instrument-biases and ionospheric delay, in the case of dual-frequency receivers.

later. The carrier-phase signal repeats after one cycle, however, and is thus ambiguous. This makes it more difficult to use carrier-phase measurements for positioning. The determination of the velocity only involves differences, i.e., the unknown number of multiples of the wavelength is eliminated. Therefore, carrier-phase measurements should be the method of choice for the determination of the velocity. For simplicity, let ϑ^k be zero-mean Gaussian:

$$\mathcal{E}[\vartheta^k] = 0, \quad \mathcal{E}[\vartheta^k \vartheta^l] = C_v^{kl}.$$

The unit vector pointing from the satellite to the receiver is very slowly time dependent, i.e., $\dot{\vec{e}}^k \simeq 0$. Therefore,

$$\frac{d}{dt}\|\vec{r} - \vec{r}^k\| = \vec{e}^k \cdot (\vec{v} - \vec{v}^k),$$

which allows determining the velocity $\vec{v}$ of the receiver and the drift $\dot{\delta}$ of his clock from:

$$\lambda\dot{\phi} = H\chi + \vartheta$$

and

$$\lambda\dot{\phi} = \begin{pmatrix} \lambda\dot{\tilde{\phi}}^1 + (\vec{e}^1)^T\vec{v}^1 + c\dot{\delta}^1 \\ \vdots \\ \lambda\dot{\tilde{\phi}}^K + (\vec{e}^K)^T\vec{v}^K + c\dot{\delta}^K \end{pmatrix}, \quad \chi = \begin{pmatrix} \vec{v} \\ c\dot{\delta} \end{pmatrix}, \quad \vartheta = \begin{pmatrix} \vartheta^1 \\ \vdots \\ \vartheta^K \end{pmatrix},$$

with λ being the wavelength, and $\tilde{\phi} \in [0,1]$ the phase of the carrier signal. The same maximum likelihood approach, as for position estimation, leads to another least squares problem with the solution

$$\hat{\chi} = (H^T V H)^{-1} H^T V \lambda \dot{\phi}.$$

In this expression V denotes the inverse correlation matrix of the noise in the derivative of the carrier-phase:

$$V = C_{\vartheta}^{-1}.$$

and H denotes the geometry matrix introduced in Equation (3.9). Since H has been previously determined, the basic equations for the determination of the velocity and clock drift are linear.

3.6 Summary

In n-dimensional space, the distances from n satellites determine the position of a receiver up to a two-fold degeneracy. In satellite navigation, distances are determined by measuring the time of arrival of radio signals, transmitted from synchronized satellites, using a local clock that is not typically not synchronized with the satellites' clocks. In this case, measurements from $n+1$ satellites are needed in order to determine the receiver's position and its clock offset. In 3 dimensions 4 satellites are thus needed. Since the equations relating the measurements to the position are non-linear and since the measurements are affected by noise, the position is determined through the solution of least-squares problems obtained by linearizing the original equations (Newton-Raphson method). Additional measurements ($n > 4$) are used in order to reduce the impact of noise. The geometry matrix H plays a crucial role in the determination of the position $\vec{r}$ and of the clock offset $c\delta$, as well as of velocity $\dot{\vec{r}}$ and of the clock drift $c\dot{\delta}$. The geometry matrix defines the dilution of precision (DOP), which describes the mentioned impact of the geometry on precision. The User Equivalent Range Error (UERE) is the other important factor in precision. It consists of a sum of individual contributions from error sources, such as orbit determination and prediction, clock noise, ionospheric and tropospheric delay, multipath propagation, as well as receiver noise.

Appendix A The Least-Squares Solution

In the present appendix, the least-squares solution associated with a maximum likelihood solution of the stochastic equation

$$\rho = H\xi + \eta,$$

is determined. The space-time vector ξ is the unknown quantity, while the geometry matrix H, and the set of measurements ρ are given. The least-squares problem consists in minimizing

$$\|\rho - H\xi\|^2_{C^{-1}} = (\rho - H\xi)^T C^{-1} (\rho - H\xi).$$

The noise η, is Gaussian with

$$\mathcal{E}[\eta] = 0, \quad \mathcal{E}[\eta\,\eta^T] = C,$$

and C is a symmetric, positive definite, and non-singular matrix.

In order to find the vector ξ, which minimizes the norm, the derivative with respect to ξ is computed. This is done using the following relations

$$\frac{\partial}{\partial \xi} a^T \xi = \begin{pmatrix} \frac{\partial}{\partial \xi_1} \\ \vdots \\ \frac{\partial}{\partial \xi_4} \end{pmatrix} \sum_{i=1}^{4} a_i \xi_i = \begin{pmatrix} a_1 \\ \vdots \\ a_4 \end{pmatrix} = a = \frac{\partial}{\partial \xi} \xi^T a,$$

with a being a constant vector, and

$$\frac{\partial}{\partial \xi} \xi^T A \xi = (A^T + A)\, \xi,$$

with A being a ξ-independent matrix. Applying these identities and using the symmetry of W, i.e. $W^T = W$ leads to:

$$\begin{aligned} \frac{\partial}{\partial \xi} \|\rho - H\xi\|_W^2 &= \frac{\partial}{\partial \xi} (\rho - H\xi)^T W (\rho - H\xi) \\ &= 2(-H^T W \rho + H^T W H \xi) \\ &= 0. \end{aligned} \tag{3.18}$$

In the case that $H^T W H$ is non-singular, Equation (3.18) provides the estimate:

$$\hat{\xi} = (H^T W H)^{-1} H^T W \rho \tag{3.19}$$

The vector of quadratic residues

$$\Delta = H\hat{\xi} - \rho$$

is typically non-zero. Its length is extremal (minimal or potentially maximal). It is easily seen that Δ is orthogonal to the image of H : let ζ be an arbitrary vector, then

$$(H\zeta)^T W \Delta = \zeta^T (H^T W H \hat{\xi} - H^T W \rho) = 0. \tag{3.20}$$

The geometric interpretation of this relation is shown in Figure 3.16. It suggests that the extremum is a minimum. This can be verified analytically by choosing an arbitrary vector ζ describing an offset from the extremum, then due to the orthogonality property from Equation (3.20):

$$\|H(\hat{\xi} + \zeta) - \rho\|_W^2 = \|H\hat{\xi} - \rho\|_W^2 + \|H\zeta\|_W^2 \geq \|H\hat{\xi} - \rho\|_W^2,$$

i.e. the extremum is a minimum.

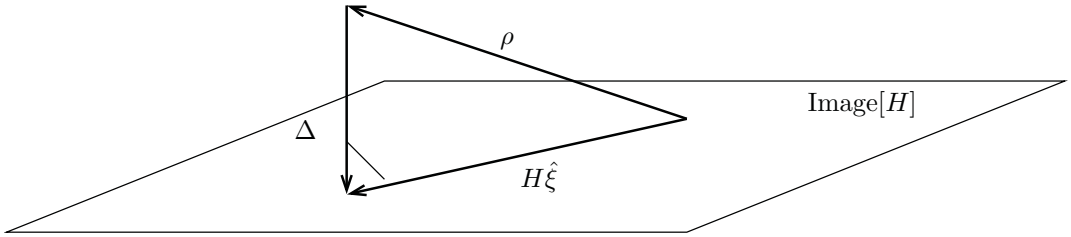

Figure 3.16: The residual error of the least-squares solution is orthogonal on the image of H.

Exercises

1. Range-based Positioning

   - Describe the solution set to range-based positioning with K satellites in K dimensions. Are there situations with a unique solution or with a continuous set of solutions? *Hint:*

Define $A^{ij} = (\vec{r}^{\,i} - \vec{r}^{\,1}) \cdot (\vec{r}^{\,j} - \vec{r}^{\,1})$, *and* n *a vector orthogonal to the span of* $\vec{r}^{\,1} \ldots \vec{r}^{\,K}$, *then prove that if the span is* $K-1$ *dimensional*

$$\vec{r} = \vec{r}^{\,1} + \sum_{j=2}^{K} \alpha^j (\vec{r}^{\,j} - \vec{r}^{\,1}) + \beta \vec{n}$$

with

$$\alpha^j = \sum_{k=2}^{K} (\vec{r} - \vec{r}^{\,1}) \cdot (\vec{r}^{\,k} - \vec{r}^{\,1}) (A^{-1})^{kj}.$$

2. Positioning in 1+1 space-time dimensions

   - Consider the situation with 1+1 space-time dimensions, as shown in Figure 3.3. Assume that the coordinates $((\vec{r}^{\,k})^T, t^k)$ of the two satellites $k = 1, 2$ are known. Determine $(\vec{r}^{\,T}, t)$. How would the solution change if all times where offset by Δt?

3. Newton-Raphson Method

   - Linearize Equation (3.6) by computing the derivative of the non-linear term $\|\vec{r} - \vec{r}^{\,k}\|$.
   - Prove the validity of Equation (3.7) by applying the Newton-Raphson Method to Equation (3.6).

4. Mathematics of least-squares

   - Let a be a vector of scalars and A a matrix of scalars, both independent of ξ, derive the identities:

     $$\frac{\partial}{\partial \xi} a^T \xi = a = \frac{\partial}{\partial \xi} \xi^T a,$$

     and

     $$\frac{\partial}{\partial \xi} \xi^T A \xi = (A^T + A)\, \xi.$$

     Hint: Use the chain rule for proving the second relation.

   - A matrix M is called positive definite iff

     $$\xi^T M \xi > 0, \quad \forall \xi \neq 0.$$

     Let η be a vector of statistically independent stochastic variables, prove that the noise covariance matrix $C = \mathcal{E}[\eta\eta^T]$ is symmetric and positive definite. Prove that this implies $\|\xi\|_C = (\xi^T C \xi)^{1/2} = 0 \Rightarrow \xi = 0$.

   - Derive the solution to the least-squares problem, i.e. Equation (3.19):

     $$\xi = (H^T W H)^{-1} H^T W \rho.$$

   - Least squares problem for velocities and clock drifts: Compute

     $$\frac{d}{dt} \|\vec{r} - \vec{r}^{\,k}\|$$

     in terms of $\vec{v} = \dot{\vec{r}}$ and $\vec{v}^{\,i} = \dot{\vec{r}}^{\,k}$.

5. Dilution of precision: Consider a two dimensional system. Assume that all clocks are perfectly synchronized, i.e. $c\delta = 0$ and $c\delta^k = 0$ for $k \in \{1, 2, \ldots, K\}$. This corresponds to a range-based positioning problem in two dimensions.

- Write the equations for the least-squares problem that must be solved to determine the position ξ from noisy measurements ρ.
- What is the minimum number of satellites K needed to compute the position of a receiver? A discrete ambiguity is admissible.
- For simplicity assume that the satellites are at unit distance. Describe a meaningful parametrization of the relative positions of the satellites.
- Compute the dilution of precision as a function of that parameter.
- Describe the geometries that lead to the smallest/largest dilution of precision. Interpret your result by making a drawing. Is there a direction in which the position estimate is particularly noisy?

Bibliography

[1] P. Hartl, "The Precise Range and Range Rate Equipement (PRARE) and its possible support to radar altimeter measurements for ERS-1," *ESA ERS-1 Radar Altimeter Data Prod.*, p. 153-160, 1984.

[2] S. Feairheller, R. Clark, "Other Satellite Navigation Systems," in *Understanding GPS, Principles and Applications,* 2nd Ed., E.D. Kaplan, C.J. Hegart (Editors), Artech House, Norwood, MA, 2006.

[3] E.W. Grafarend, J. Shan, "A Closed-Form Soluation of the Nonlinear Pseudo-Ranging Equations (GPS)," *Planetary Geodesy,* vol. 28, no. 3, (Articical Satellites), 1996.

[4] P. Jurkowski, P. Henkel, G. Gao, C. Günther, "Integer Ambiguity Resolution with Tight and Soft Baseline Constraints for Freight Stabilization at Helicopters and Cranes," *Proc. ION-ITM '11,* pp. 336-346, Jan. 2011.

[5] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory,* Prentice Hall, Upper Saddle River, NJ, 1993

[6] B. Belabbas, *Private Communication,* 2009.

[7] B. Belabbas, F. Petitprez, A. Hornbostel, "UERE Analysis for Static Single Frequency Positioning Using Data of IGS Stations," *Proc. NTM 2005*, pp. 310-319, Jan. 2005

[8] K. Kovach, "New User Equivalent Range Error (UERE) Budget for the Modernized Navstar Global Positioning System (GPS)", *Proc. NTM 2000,* pp. 550-573, Jan. 2000

[9] FAA, `http://www.nstb.tc.faa.gov/reports/pan58\_0707.pdf`

[10] P. Henkel, Z. Wen, C. Günther, "Estimation of satellite and receiver biases on multiple Galileo frequencies with a Kalman filter," *Proc. ION ITM,* pp. 1067-1074, 2010.

[11] K. Kovach, "SVN-23/PRN-23 Integrity Failure of 1st. January 2004" `http://www.navcen.uscg.gov/cgsic/meetings/43rdmeeting/09%20PRN-23%2043.PPT`

[12] RTCA/DO-229C, *Minimum Operational Performance Standards for Global Positioning System/Wide Area Augmentation System Airborne Equipment*, 2001