

- 一、因子图优化的优势
- 二、因子图状态估计方法
 - 1、因子图模型
 - 2、GNSS中的因子图模型
 - 3、因子图算法求解
- 三、基于FGO的伪距单点定位算法
- 四、基于FGO的载波相位差分定位算法

一、因子图优化的优势

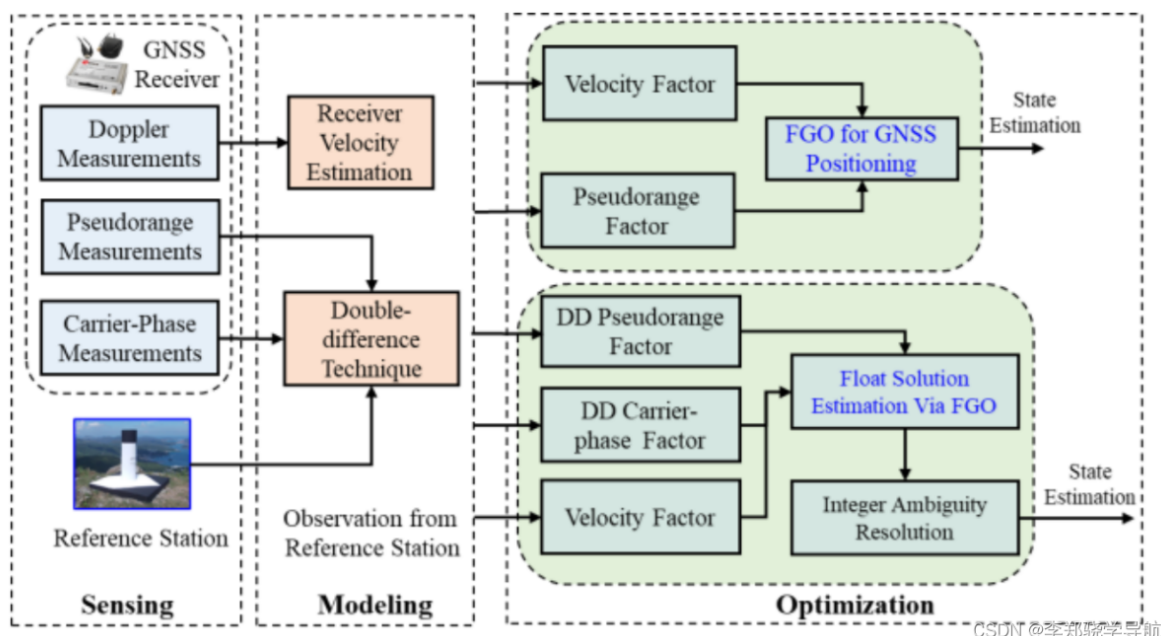
GNSS在诸如城市、施工现场、农田、密林等城市峡谷等遮挡环境下：

- 观测值数量和质量不稳定，卫星数甚至不足以EKF定位。
- 伪距受**多路径**影响严重，载波相位容易产生**周跳**。
- **非视距信号**（NOLS）增多，且基于传统的接收机技术很难完全检测出来。
- 卫星的**几何分布差**，即使我们检测出来了潜在的非视距信号，然后将其排除，这会导致更差的卫星几何分布。

且由于在此类环境下，这些测量值的**时间相关性**更加明显，使用传统的滤波手段进行处理仅考虑前一状态的值忽略了其他历史状态值，使得定位结果更加不精确。为充分挖掘历史信息的作用，引入**因子图优化（FGO）**对GNSS测量值进行融合解算，该方法**将伪距、载波相位、多普勒观测量添加为因子节点，引入多普勒因子连接历元间变量节点**，实现解算模型的因子图优化。相较于滤波，因子图的优势体现在：

- 能够进行多次线性化计算以及**多次迭代**以便更好的逼近真实的模型。
- **充分利用历史观测信息**，能够将所有的历史信息加入到对当前状态的计算，提高系统的精度
- 对于某些具有**时间相关性**的测量值更为有效，比如GNSS数据在**相邻多历元之间强相关**，钟的漂移在多帧内是不怎么变的，载波观测量具备类似的整周模糊度。
- 不需要设计复杂的设计矩阵，且可以**灵活的增加约束关系**。
- EKF有很多变种，IEKF具备因子图优化的多次迭代，如果考虑多窗口的EKF的话，就会也具备多帧数据的考量了，这样就和因子图很接近了。但**因子图很灵活**，可以随意变化窗口数目，不需要复杂的重新设计观测矩阵之类的，最关键是可以很灵活的增加一些新的约束关系。
- EKF的一阶马尔科夫模型假设，前提是观测量的协方差可以准确的计算出来并且应用到EKF中，但是这个是在实际中，特别是都市环境下很难的，GNSS的观测量不确定性很大。

因子图技术的潜力，也在一个近期的Google在美国导航协会的ION GNSS+ 2021的会议上举办的一个手机GNSS定位的比赛上获得一个极大的验证，来自日本千叶工业大学的Taro Suzuki博士使用因子图技术，对手机采集的GNSS数据进行定位解算，获得了第一名。



二、因子图状态估计方法

1、因子图模型

因子图作为一种数学工具，是用因子描述多变量复杂函数的二部图，通常被用于SLAM中。**图优化**就是将优化问题转换成图的形式，图由**边**和**顶点**组成。**边**连接着**顶点**，表示**顶点之间的一种关系**。顶点表 估计问题中的未知随机变量，而因子表示有关这些变量的概 率信息，这些信息是从测量或先验信息中得出的。

因子图是一种二分图模型，它表征了**全局函数和局部函数之间的关系**，同时也表示了**各个变量与局部函数**之间的关系，以下式为例：

$$g(x_1, x_2, x_3, x_4, x_5) = f_A(x_1)f_B(x_2)f_C(x_1x_2x_3)f_D(x_3x_4)f_E(x_3x_5)$$

将全局函数 g 转化为局部函数 f 的乘积，其对应的因子图如下所示：

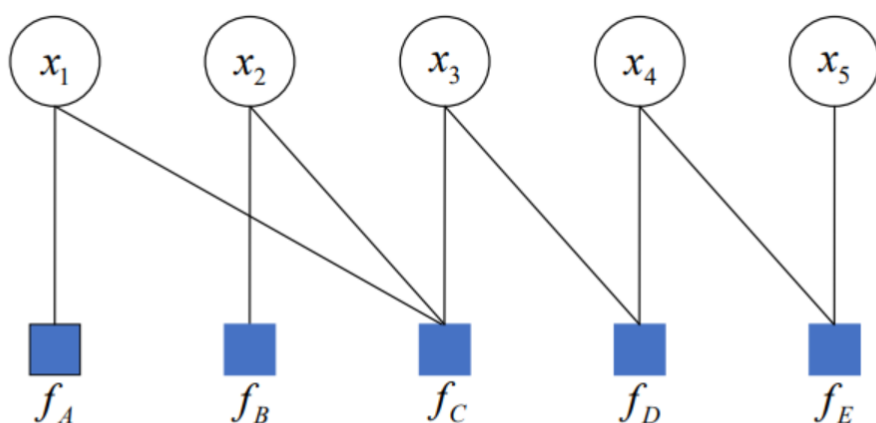


Figure 1. The example of expression of factor graph

$$g(x_1, x_2, x_3, x_4, x_5)$$

图 1 $g(x_1, x_2, x_3, x_4, x_5)$ 的因子图表示示例

- 圆圈：每一个变量 x 对应一个**变量结点**
- 正方形：每一个局部函数 f 对应一个**因子结点**

- 线：当且仅当变量 x 是局部函数 f 的自变量时，相应的变量节点和因子节点之间有一条边连接两个节点

2、GNSS中的因子图模型

在以往的状态估计问题中，往往是寻找 X 来最好地描述观测值 Z 。根据**贝叶斯法则**，状态量 X 和观测值 Z 的**联合概率等于条件概率乘以边缘概率**：

$$P(X, Z) = P(Z | X)P(X)$$

式中： $P(Z|X)$ 为观测值 Z 对应的概率； $P(X)$ 是状态量 X 的先验概率。后验分布 $P(X|Z)$ 是一种常用且直观评估状态集和观测集之间拟合程度的方法，我们求解期望的状态集可以由通过后验分布的最大化来实现，也就是**极大后验估计**：

$$\hat{X} = \arg \max_X P(X | Z)$$

基于因子图的状态估计方法正是将状态估计理解为**对系统联合概率密度函数的极大后验估计问题**。一个系统可以描述为**状态方程**和**量测方程**两部分，并将状态误差和量测误差视为**零均值白噪声**即：

$$\begin{cases} x_k = f_k(x_{k-1}, u_k) + w_k, & w_k \sim N(0, \Sigma_k) \\ z_k = h_k(x_k) + v_k, & v_k \sim N(0, \Lambda_k) \end{cases}$$

根据正态分布的特性可以得到真实状态 k_x 和理想量测 k_z 的条件概率分布满足：

$$\begin{cases} P(x_k | x_{k-1}) \propto e^{-\frac{1}{2} \|f_k(x_{k-1}) - x_k\|_{\Sigma_k}^2} \\ P(z_k | x_k) \propto e^{-\frac{1}{2} \|f_k(x_k) - z_k\|_{\Lambda_k}^2} \end{cases}$$

实际中的状态量 X 往往是不知道的，而当前状态下的观测 Z 是知道的，也就是 $P(Z|X)$ 是知道的，因此在因子图模型中：

$$X_k^* = \arg \max P(X_k | Z_k) \propto \arg \max P(X_k)P(Z_k | X_k)$$

其中， $X_k = \{x_{0:k}\}$ 是状态的集合， $Z_k = \{z_{0:k}^j\}$ 是所有状态下量测的集合。若系统服从马尔科夫假设，那么：

$$X_k^* = \arg \max_{X_k} P(X_k | Z_k) \propto \arg \max_k P(x_0) \prod_k \left[P(x_i | x_{i-1}) \prod_{m_i}^{m_i} \left[P(z_i^j | x_i) \right] \right]$$

对式取对数得到后，将式代入式可以得到，系统的状态估计可等价于全局损失函数的联合优化：

$$X^* = \arg \min_X \sum_i^k \left\{ \|f_i(x_{i-1}, u_i) - x_i\|_{\Sigma_i}^2 + \sum_{j=1}^{m_j} \|h_i^j(x_i) - z_i^j\|_{\Lambda_{ij}}^2 \right\}$$

上式即为基于因子图优化的估计的一般表达式，其左项为系统状态转移过程，右项为量测过程， Σ 和 Λ 分别是状态转移过程和量测过程的协方差矩阵，进行求解的是状态集合 x 。对于式，可以用下图进行表示：

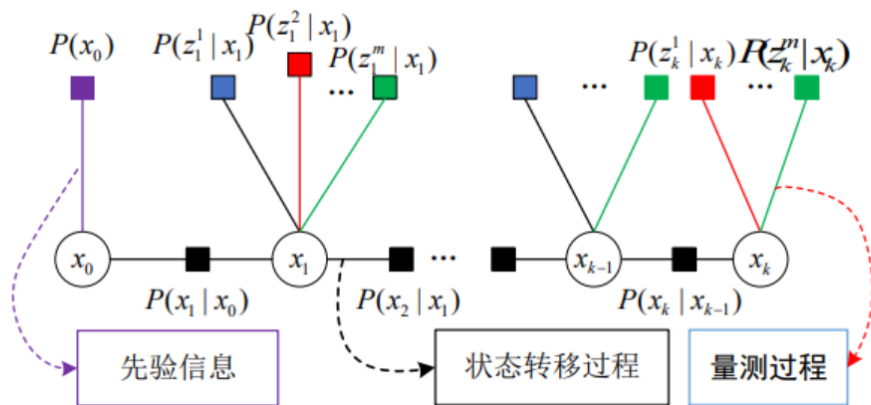


Figure 2. Representation of joint optimization of global cost functions with factor graph

图 2 全局代价函数的联合优化的因子图表示

CSDN @李郑晓学导航

- 圆圈：变量节点，表示系统待估计的状态，对应一个变量 x
- 正方形：因子节点，表示先验信息、状态转移和量测过程，对应一个局部函数 f ，其中：
 - 紫色 $P(x_0)$ 为先验因子
 - 黑色 $P(x_1|x_0) \dots P(x_k|x_{k-1})$ 为状态转移，由上一时刻状态推测下一时刻状态
 - 其它为量测信息， $P(z|x)$ 表示在参数 x 的条件下得到观测值 z
- 线：当且仅当变量 x 是局部函数 f 的自变量时，相应的变量节点和因子节点之间有一条边连接两个节点
- 若在模型中加入其他传感器，只需将其添加到框架中相关的因子节点处即可

利用因子图模型对估计系统的联合概率密度函数进行表示，可以直观地反映动态系统的动态演化过程和每个状态对应的量测过程。同时，图形化的表示使系统具有更好的通用性和扩展性。

3、因子图算法求解

每一个观测变量在上面贝叶斯网络里都是单独求解的（相互独立），所以所有的条件概率都是乘积的形式，且可分解，在因子图里面，分解的每一个项就是一个因子，乘积乘在一起用图的形式来描述就是因子图。整个因子图实际上就是每个因子单独的乘积。求解因子图就是将这些因子乘起来，求一个最大值，得到的系统状态就是概率上最可能的系统状态。

先找到残差函数 $e(x)$ ，由因子节点可以得到我们估计的值和实际的测量值之间的差值，即每个因子 f 会对应一个残差函数。根据中心极限定理，绝大多数传感器的噪声是符合高斯分布的，所以每个因子都是用高斯分布的指数函数来定义的。

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

指数函数对应了残差函数，包括两个部分：系统状态量和观测量。残差函数实际上表示的是用状态量去推测的观测量与实际观测量的区别。残差函数的表达式一般都是非线性的，可以通过改变变量 X 来使残差函数最小化，残差函数最小，估计的值越符合观测值，套到因子图里面来看，因子图的求解是要所有因子的乘积最大化，

$$\hat{X} = \arg \max_X \prod_i \exp\left(-\frac{1}{2}|e_i(X_i)|_{\Sigma_i}^2\right)$$

对于负指数函数形式，每一个因子乘积最大化代表里面的 $e(x)$ 最小化，对目标函数取对数，概论问题转为非线性最小二乘问题：

$$\hat{X} = \arg \max_X \sum_i (e_i(x_i))^2$$

非线性最小二乘可以选择**高斯-牛顿法**、**列文伯格-马夸尔特**或者**Dogleg**等迭代优化算法求解最优值，高斯-牛顿法比较简单，但稳定性较差，算法可能不收敛；列文伯格-马夸尔特引入**置信区间**概念，约束下降的步长，提高稳定性，Dogleg也类似。**问题性质较好时可用高斯-牛顿法，问题条件恶劣时选择列文伯格-马夸尔特或者Dogleg。**

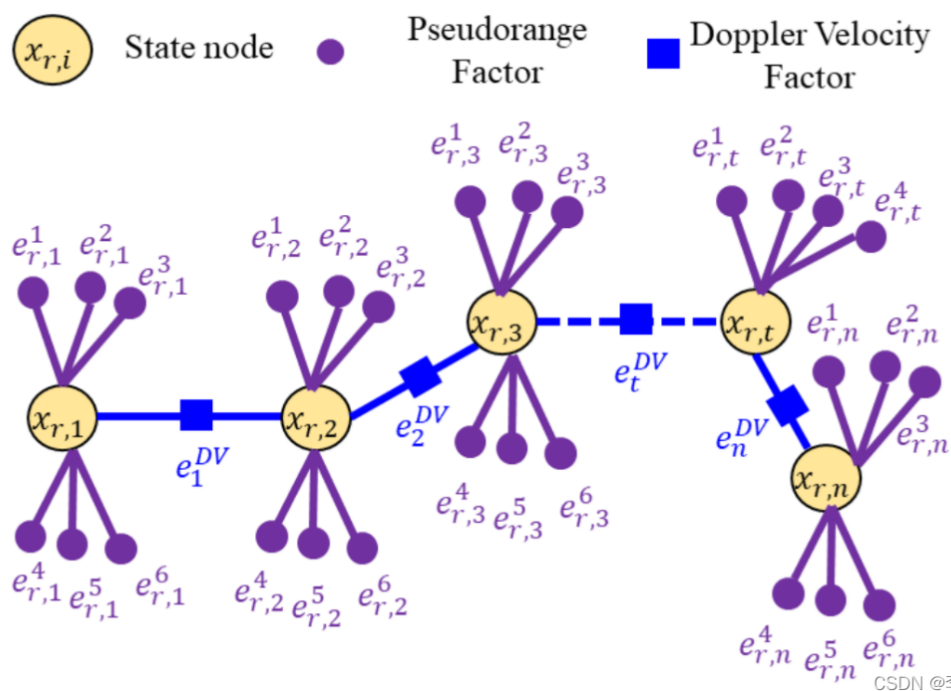
几种非线性最小二乘解法比较：

最速下降法：目标函数在 x_k 处泰勒展开，保留一阶项， $x^* = -J(x_k)$ ，最速下降法过于贪心，容易走出锯齿路线，反而增加迭代次数。**牛顿法**：二阶泰勒展开，利用二次函数近似原函数。 $H * X = -J$ ，牛顿法需要计算目标函数的海森矩阵阵，计算量大。规模较大时比较困难。**高斯-牛顿法 (GN)**： $f(x)$ 进行一阶泰勒展开， $f(x)$ 而不是 $F(x)$ ，高斯牛顿法用雅各比矩阵 JJ^T 来作为牛顿法中二阶海森阵的近似， $HX = g$ ，在使用高斯牛顿法时，可能出现 JJ^T 为奇异矩阵或者病态的情况，此时增量稳定性较差，导致算法不收敛。**列文伯格-马夸尔特方法 (LM)**：基于信赖区域理论，是由于高斯-牛顿方法在计算时需要保证矩阵的正定性，于是引入了一个约束，从而保证计算方法更具普适性。 $(H + \lambda I)x = g$ ，当 λ 较小时， H 占主导，近似于高斯牛顿法，较大时， $\lambda * I$ 占主导，接近最速下降法。

三、基于FGO的伪距单点定位算法

基于因子图的伪距单点定位算法使用伪距、多普勒融合算法选取的状态变量为：

$X_n = [x_n, y_n, z_n, \delta t_n]^T$ ， $p_{u,t} = [x_t, y_t, z_t]^T$ ，是某时刻 t 的位置，状态变量集合为 $\chi = [X_1, X_2, \dots, X_n]$ 。选取**伪距**作为测量值，**多普勒**的测速结果作为状态转移的约束。如下图：



CSDN @李郑晓学导航

伪距的观测模型为：

$$\rho_{u,t}^s = r_{u,t}^s + c \left(\delta t_{u,t} - \delta t_{u,t}^{(s)} \right) + I_{u,t}^s + T_{u,t}^s + \varepsilon_{u,t}^s$$

其中， $r_{u,t}^s$ 指 t 时刻的卫星 s 与接收机的距离， $\delta t_{u,t}^{(s)}$ 、 $I_{u,t}^s$ 、 $T_{u,t}^s$ 指卫星钟差、电离层误差、对流层误差，可以通过假定的模型进行估算， $\varepsilon_{u,t}^s$ 指随机误差。经过误差校正后的**伪距观测模型**为：

$$\rho_c = f_c(p_{u,t}, p_t^s, \delta_{u,t}) + \omega_{\rho,t}$$

其中, $f_c(p_{u,t}, p_t^s, \delta_{u,t}) = \|p_{u,t} - p_t^s\| + \delta_{u,t}$, p_t^s 是卫星位置, $\omega_{\rho,t}$ 是伪距的测量误差, 其协方差矩阵为 $\Sigma_{u,t}^s$, 可以通过高度角模型、信号质量等建立起协方差矩阵。那么, 伪距测量的损失函数为:

$$\|e_{u,t}^s\|_{\Sigma_{u,t}^s}^2 = \|\rho_c - f_c(p_{u,t}, p_t^s, \delta_{r,t})\|_{\Sigma_{u,t}^s}^2$$

目前, GNSS的多普勒测速精度能够达到0.03m/s, 远远高于伪距的精度, 多普勒定速的精度影响可以忽略不计, 利用多普勒测速结果对相邻历元进行约束。多普勒测速的观测模型为:

$$v_{u,t} = f_{v,t}(p_{u,t+1}, p_{u,t}) + \omega_{v,t}$$

其中, $f_{v,t}(p_{u,t+1}, p_{u,t}) = (p_{u,t+1} - p_{u,t})/\Delta t$, 速度 = 两时刻的位置差值/时间间隔, Δt 是两状态的时间间隔, 即接收机接收数据频率的倒数, 可以预见的是数据接收频率越高, 模型越接近实际。 $\omega_{v,t}$ 是多普勒测速的测量误差, 其协方差为 $\Sigma_{v,t}$, 可以通过接收机的标称测速精度或者多次实验进行调整。那么, 多普勒测速的损失函数为:

$$\|e_{v,t}\|_{\Sigma_{v,t}}^2 = \|v_{u,t} - f_{v,t}(p_{u,t+1}, p_{u,t})\|_{\Sigma_{v,t}}^2$$

可以得到基于因子图优化的融合伪距与多普勒测量值的定位解算方法的目标函数是:

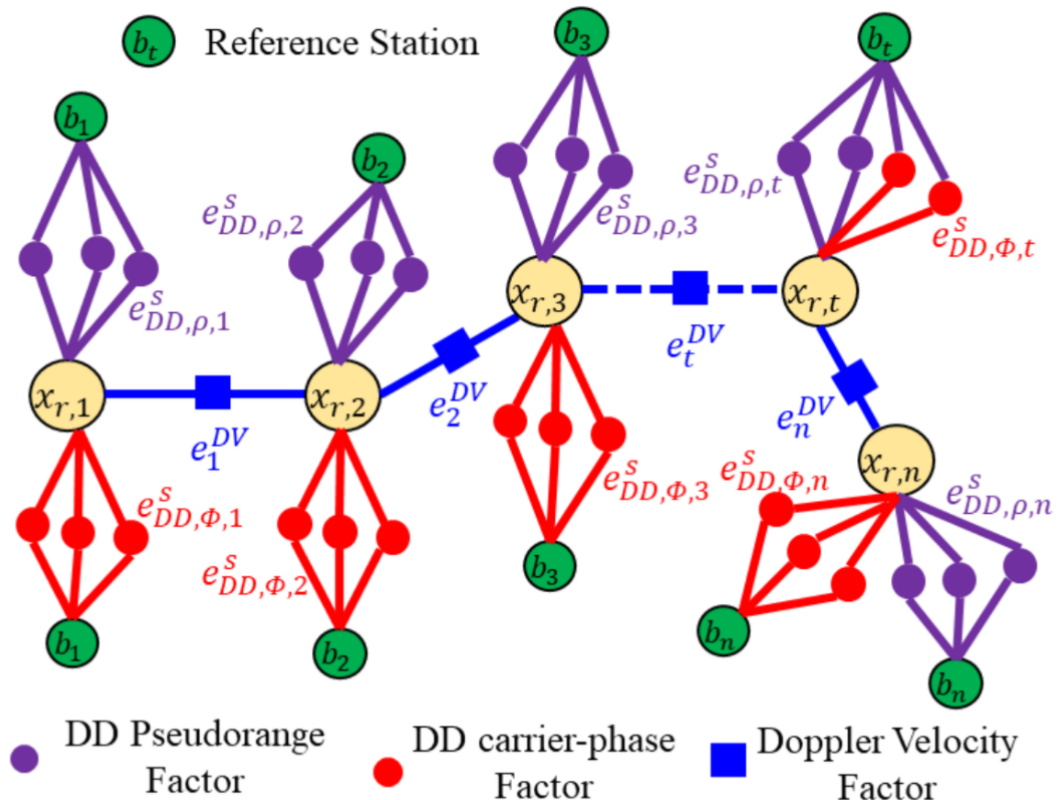
$$\chi^* = \arg \min_{\chi} \sum_t \|e_{v,t}\|_{\Sigma_{v,t}}^2 + \|e_{u,t}^s\|_{\Sigma_{u,t}^s}^2$$

其中, 需要进行最优化计算的 χ 是状态的集合, 而非单独的历元。

四、基于FGO的载波相位差分定位算法

基于因子图的载波相位差分定位算法使用伪距、载波相位、多普勒进行融合算法, 选取的状态变量

$X_t = [x_t, y_t, z_t, \Delta \nabla N_{ur,t}^1, \dots, \Delta \nabla N_{ur,t}^{m-1}]^T$, $p_{u,t} = [x_t, y_t, z_t]$ 是移动站的位置, $[\Delta \nabla N_{ur,t}^1, \dots, \Delta \nabla N_{ur,t}^{m-1}]^T$ 是 t 时刻下移动站 u 和基准站 r 之间 m 颗共视卫星的 $m-1$ 个双差整周模糊度, 状态变量的集合为 $\chi = [X_1, X_2, \dots, X_n]$ 。选取伪距双差、载波相位双差作为测量值, 多普勒的测速结果作为状态转移的约束。如下图:



对基准站和移动站作站间差、星间差后，可以得到**伪距双差的测量模型**为：

$$\rho_{DD,\rho,t}^s = f_{DD,\rho,t}^s(p_{u,t}, p_{b,t}, p_t^s, p_t^b) + \omega_{DD,\rho,t}^s$$

其中， $f_{DD,\rho,t}^s(p_{u,t}, p_{b,t}, p_t^s, p_t^b) = \|p_{u,t} - p_t^s\| - \|p_{r,t} - p_t^s\| - \|p_{u,t} - p_t^b\| + \|p_{r,t} - p_t^b\|$ ， p_t^b 是共视卫星中的参考星位置，为了使双差测量模型更加精确，一般选取共视卫星中高度角最大的一个作为参考星。 $\omega_{DD,\rho,t}^s$ 是伪距双差测量模型的测量误差，是由于即使经过双差处理，若基准站、移动站的大气情况相差过大，该测量模型仍存在误差，其协方差为 $\Sigma_{DD,\rho,t}^s$ ，那么，**伪距双差测量损失函数**为：

$$\|e_{DD,\rho,t}^s\|_{\Sigma_{DD,\rho,t}^s}^2 = \|\rho_{DD,\rho,t}^s - f_{DD,\rho,t}^s(p_{u,t}, p_{b,t}, p_t^s, p_t^b)\|_{\Sigma_{DD,\rho,t}^s}^2$$

同理，**载波相位双差的观测模型**表示为：

$$\phi_{DD,\rho,t}^s = f_{DD,\phi,t}^s(p_{u,t}, p_{b,t}, p_t^s, p_t^b, \Delta\nabla N_{ur,t}^1, \Delta\nabla N_{ur,t}^2, \dots, \Delta\nabla N_{ur,t}^{m-1}) + \omega_{DD,\phi,t}^s$$

其中，

$f_{DD,\phi,t}^s(p_{u,t}, p_{b,t}, p_t^s, p_t^b) = (\|p_{u,t} - p_t^s\| - \|p_{r,t} - p_t^s\|) - (\|p_{u,t} - p_t^b\| - \|p_{r,t} - p_t^b\|) + \Delta\nabla N_{ur,t}^s$ ， $\omega_{DD,\phi,t}^s$ 是载波相位双差的测量误差，协方差为 $\Sigma_{DD,\phi,t}^s$ ，**载波相位双差损失函数**为：

$$\|e_{DD,\phi,t}^s\|_{\Sigma_{DD,\phi,t}^s}^2 = \|\phi_{DD,\phi,t}^s - f_{DD,\phi,t}^s(p_{u,t}, p_{b,t}, p_t^s, p_t^b)\|_{\Sigma_{DD,\phi,t}^s}^2$$

多普勒测速的损失函数同伪距单点定位的式子。可以得到基于因子图优化的融合伪距、载波相位和多普勒测量值的定位解算方法的**目标函数**：

$$\chi^* = \arg \min_{\chi} \sum_t \|e_{v,t}\|_{\Sigma_v}^2 + \|e_{DD,\phi,t}^s\|_{\Sigma_{DD,\phi,t}^s}^2 + \|e_{DD,\rho,t}^s\|_{\Sigma_{DD,\rho,t}^s}^2$$

其中，**需要进行最优化计算的 χ 是状态的集合**，而非单独的历元。通过对目标函数的求解即可得到各个历元的浮点解以及双差整周模糊度的浮点解，**随后进行整周模糊度固定**。如果能够固定再次进行计算即可得到精确的定位解，如果不能固定即用浮点解作为最终的解算结果。