EDGELIST PHASE UNWRAPPING ALGORITHM FOR TIME SERIES INSAR ANALYSIS

Piyush Agram,1,* and Howard Zebker1
1Departments of Electrical Engineering and Geophysics, Stanford University, 350 Serra Mall, Packard 334, Stanford, CA 94305, USA
*Corresponding author: shanker@stanford.edu

We present here a new integer programming formulation for phase unwrapping of multi-dimensional data. Phase unwrapping is a key problem in many coherent imaging systems, including time series synthetic aperture radar interferometry (InSAR), with two spatial and one temporal data dimensions. The minimum cost flow (MCF) [1] phase unwrapping algorithm describes a global cost minimization problem involving flow between phase residues computed over closed loops. We replace closed loops by reliable edges as the basic construct, thus leading to the name “edgelist”. Our algorithm has several advantages over current methods – it simplifies the representation of multi-dimensional phase unwrapping, it allows for incorporation of data from external sources, like GPS, where available to better constrain the unwrapped solution, and it treats regularly sampled or sparsely sampled data alike. It thus is particularly applicable to time series InSAR where data are often irregularly spaced in time and individual interferograms can be corrupted by large decorrelated regions. Similar to the MCF network problem, our formulation also exhibits total unimodularity (TUM) which enables us to solve the integer program using efficient linear programming tools [11]. We apply our method to a PS-InSAR dataset from the creeping section of the Central San Andreas Fault. The validity of our approach has been confirmed by the independent development of similar techniques by Costantini et al [12].

1. Introduction

Phase unwrapping as used in InSAR geodesy is the reconstruction of absolute phase from measured phase known only modulo 2π on a finite grid of points. Interferograms, especially in time series analysis, are often irregularly sampled in both space and time, so that existing algorithms do not always properly unwrap the data. In this paper, we develop a method to address the most general form of the phase unwrapping problem – unwrapping sparsely distributed, multi-dimensional wrapped phase data. Since we focus on applying our new technique to InSAR phase data, we restrict our discussion to conventional two dimensional (single interferogram) InSAR data and multi-temporal InSAR (Persistent Scatterer and Short Baseline methods) three dimensional data sets. The performance of any phase unwrapping algorithm depends on our ability to estimate the phase gradient between two adjacent samples in our data set. The basic assumption in all phase unwrapping problems is that the underlying continuous unwrapped phase function \( \phi \) is well sampled in every dimension to enable us to reconstruct it from wrapped phase measurements \( \psi \) except at a finite, relatively small, number of discontinuities. The existence of discontinuities in a data set produces path dependent inconsistencies or residues [3]. The branch cut algorithm [3] and its derivatives are very popular approaches to phase unwrapping, due to their ease of implementation. Hooper and Zebker [4] successfully extended the idea of the shortest branch cut to the 3D phase unwrapping problem.

Costantini [1] developed the first network programming formulation to solve the regularly sampled two dimensional (2D) phase unwrapping problem using an \( L_1 \) norm minimum cost flow (MCF) approach. Costantini and Rosen [2] adapted the algorithm to solve the irregularly sampled 2D phase unwrapping problems using Delaunay triangulation. The performance of the network programming based unwrapping algorithms for 2D data sets was further improved by the development of an iterative \( L_0 \) norm approximation algorithm [5] and its application in combination with statistical cost functions [6].

In this paper, we propose a new minimum \( L_1 \) norm formulation that is more flexible than the original MCF formulation, allowing us to impose additional constraints on the unwrapped solutions based on available \textit{a priori} information. It also can be easily extended to analyze multi-dimensional data sets. We describe our new algorithm in Section 2. In Section 3, we explain how we can incorporate additional geodetic observations as constraints. We apply the method to a time series of data acquired over the creeping section of the San Andreas Fault in Section 4. Finally, in Section 5, we discuss implementation, draw some conclusions and suggest possible improvements.

2. Phase unwrapping formulation

Let \( V \) represent the set of points \( \{ V \mid N \} \) on which the set of measured wrapped phase values, \( \{ \psi, \phi \} \) where \( i \in \{ 1, \ldots, N \} \), is defined. Let \( E \) define the set of reliable edges \( \{ E \mid M \} \) on \( V \) such that, for every edge \( (i, j) \in E, \ i < j \). Thus \( G := \{ V, E \} \) represents a...
directed graph which will be used to estimate the set of unwrapped phase values, \( \{ \phi_i \} \) where \( i \in \{1, \ldots, N\} \). For simplicity, we define a new variable \( \pi \) where

\[
\phi_i = \psi_i + 2\pi n_i \quad \text{where} \quad i \in \{1, \ldots, N\}
\]

(1)

where \( n_i \) represent the integer number of cycles that must be added to each point of the wrapped function to obtain the unwrapped function. These variables can be interpreted as node potentials [7]. For every edge \((i, j)\) in \(E\), we define a new variable \( K_{ij} \) such that

\[
n_j - n_i + K_{ij} = \left\lfloor \frac{\psi_j - \psi_i}{2\pi} \right\rfloor \quad \text{where} \quad (i, j) \in E
\]

(2)

Here \([\cdot]\) represents the nearest integer function and \( K_{ij} \) represents the integer flow along the directed edge \((i, j)\) and is equivalent to variables \( K_1 \) and \( K_2 \) in [1]. As in the MCF formulation, we associate a non-negative convex cost multiplier \( C_{ij} \) with the integral flow on every edge \((i, j) \in E\). Also, for an L1 norm implementation we represent the flow \( K_{ij} \) as a difference of two non-negative integer flow variables \( P_{ij} \) and \( Q_{ij} \). We can then state our minimum L1 norm cost flow problem as

\[
\text{Minimize} \quad \sum_{(i, j) \in E} C_{ij} (P_{ij} + Q_{ij})
\]

(3)

subject to

\[
n_j - n_i + P_{ij} - Q_{ij} = \left\lfloor \frac{\psi_j - \psi_i}{2\pi} \right\rfloor \quad \forall (i, j) \in E
\]

(4)

\[
n_i \in \text{integer}, \quad i \in \{1, \ldots, N\}
\]

(5)

\[
P_{ij}, Q_{ij} \in \text{positive integers}, \quad (i, j) \in E
\]

(6)

Our formulation differs from that of [1] in that the basic unit in our algorithm is an edge on the unwrapping grid instead of a closed loop, hence the name “edgelist”. The salient features of the edgelist formulation include:

- The edgelist formulation reduces to the original MCF formulation if every edge of the Delaunay tessellation is included as a constraint (See Fig 1). The constraint of our phase unwrapping formulation when applied to the edges of a loop produces a loop constraint in the original MCF formulation.

- The edgelist formulation does not distinguish between 2D and 3D data, where the third dimension is generally the time dimension in a series of interferograms. Each phase measurement is treated as a distinct vertex of the graph \(G\). Consequently, more variables and constraints are needed to completely define a problem.

- The constraint matrix of the edgelist formulation is a total unimodular matrix (TUM) [11] and the right hand side of the constraints in Equation 4 is an array of integers. Similar to the MCF and other TUM integer programming problems, the edgelist formulation can also be exactly solved as a linear program (LP) when the associated objective functions minimize the \( L_1 \) norm [11].

- The edgelist formulation can readily incorporate other geodetic measurements such as GPS or leveling data as additional constraints without affecting the TUM property of the constraint matrix. This is discussed in detail in section 3.
mode statistical cost functions developed by Chen and Zebker [6] here.

3. Incorporating other geodetic measurements as constraints

Often, complementary geodetic measurements such as GPS networks or leveling surveys are available in addition to frequent SAR acquisitions over regions of interest. This information can be used to direct edges in branch cut based algorithms or adjust cost functions in a network programming method for unwrapping interferometric phase, constraining the result to reflect these additional data. In the original MCF formulation, constraints based on additional observations can be defined but at the cost of violating the TUM property of the constraint matrix [11]. Violation of the TUM property renders the problem unsolvable exactly using LP solvers.

In the case of the edgelist formulation, if alternate geodetic measurements are available at points \( p \) and \( q \), we introduce a new edge between the points (see Figure 2) and a new constraint in the formulation

\[
n_p - n_q = \Delta N_{p,q} - \left\lfloor \frac{\phi_p - \phi_q}{2\pi} \right\rfloor \tag{7}
\]

where \( \Delta N_{p,q} \) is the expected number of unwrapped phase cycles between \( p \) and \( q \). The new constraint adds additional entries to the original node but retains the TUM structure [11]. If the points \( p \) and \( q \) were already connected in the original unwrapping grid \( G \), the corresponding constraint for the edge in the unconstrained formulation (Equation 4) can be replaced by the new equation (Equation 7).

Independent geodetic estimates of line of sight displacement for \( L \) vertices in \( V \) will allow us to construct \( C(L,2) = \binom{L}{2} \) additional equations to constrain our unwrapped solutions. Simpler upper bound constraints of the form

\[
0 \leq P_{ij} \leq u \quad \forall (i,j) \in E \tag{8}
\]

\[
0 \leq Q_{ij} \leq u \quad \forall (i,j) \in E \tag{9}
\]

where \( u \) is a positive integer, are further applied to reduce the solution space. These constraints do not affect the TUM property [11]. The primary advantage of the edgelist formulation is that it provides us with controls over every data point \( (n_i) \) and every edge \( (P_{ij}, Q_{ij}) \), as opposed to control over the edges alone in case of the MCF formulation. Both these properties can be suitably exploited to solve challenging unwrapping problems as shown in the next section.

4. Case study: Creeping section of the Central San Andreas Fault

We applied our edgelist unwrapping algorithm to a persistent scatterer InSAR (PS-InSAR) data set covering an area of 40 km x 40 km around the Monarch and Austin peaks (Fig 4a) in the Central San Andreas Fault region. InSAR stacking in earlier studies characterized the spatial variation in slip deficit on the Central San Andreas Fault [10]. The presence of large decorrelated areas close to the fault severely compromised the ability to reliably estimate the deformation just north of the fault.

We processed 21 SAR scenes (Track 27, Frame 2781) acquired by ERS-1 and ERS-2 satellites between 1992 and 2004. We selected a scene from March 1997 as the master scene from minimization of the perpendicular baseline and the temporal baseline, and generated 20 interferograms. To optimize the accuracy of the correlated phase estimates in the sparse PS network, we limited the maximum perpendicular baseline to 400 m. We applied the maximum likelihood persistent scatterer (MLPS) selection algorithm [8] and found a sparse network of 2067 PS points per interferogram. Although the MLPS algorithm identified more PS points than other public domain algorithms, the PS density (1 per sq km) is still very low compared to the suggested threshold of 4 PS / sq km [9] recommended for conventional PS-InSAR phase unwrapping algorithms.

Fig 4(a) shows the average LOS displacement rate in mm/yr and Fig 4(b) shows the average profile computed as a function of distance from the fault as computed by our algorithm (Equations 3-6, 8-9 above).

To form this estimate, we also

1. Neglected elevation variation in the atmospheric phase screen, as the topography does not change by more than a few hundred feet in this area.
2. Introduced an interferogram consisting purely of zero interferometric phase values to represent the combination of the master scene with itself, into the time-series. And the node potential variables \( (n_i) \) for the vertices on the zero interferogram were constrained to be zero (See Fig 3a). This establishes a zero-reference frame with no discontinuities across the fault.
3. Chose the PS point with the highest temporal coherence as the reference PS point (see Fig 3a). The node potential for the master PS point is also constrained to be zero in all the interferograms. Thus, all the unwrapped phase values are estimated with reference to the master PS point.
4. The cost functions associated with edges of the Delaunay triangulation and all time-edges of the points that lie within a 1km zone around the fault were decreased. (See Fig 3b) This incorporates our knowledge that most of the creep is near the
surface and that there is less deformation further away from the fault. This is a good assumption for this fault but may not pertain to other parts of the world.

One of the main advantages of our method is that we do not require a temporal model for estimating deformation. The approach described above models the points on opposite sides of the fault as connected by loose (subsidized costs) strings, i.e., the unwrapped solution for a pixel depends more on the phase values of pixels on the same side of the fault as itself and less on the phase of pixels on the other side of the fault. Phase cycle jumps are preferentially compensated in the buffer zone. As a consequence, the unwrapped solutions in the buffer zone may not be entirely reliable due to the artificial discounts applied on the cost functions. But as we move away from the buffer zone, the solutions should be more accurate. Fig 4 shows the results obtained using our new algorithm.

Conventional PS-InSAR phase unwrapping algorithms failed to accommodate multiple cycle jumps across the fault for the large temporal baseline interferograms in the data set. We overcome this problem by subsidizing the costs for edges cutting across the fault. In this area the fault creeps at roughly 22 mm/yr (approximately 7 mm/yr in the radar line of sight, LOS), so that interferograms with temporal separation of more than 4 years exhibit multiple cycles across the fault. In an $L_1$ or $L_2$ norm formulation, multiple cycle jumps are penalized more heavily than single cycle jumps and require careful, and often impractical, adjustment of the cost functions. Also, the San Andreas Fault fully bisects the image and forms a line discontinuity. In other words, there is no direct connection between the regions on either side of the fault through an area that is not noisy in phase. Hence, it is not possible to unwrap around the fault in this data set and correct for the phase jumps across the fault using conventional algorithms.

5. Discussion and conclusions

We propose a new phase unwrapping algorithm that is as accurate but more flexible than previously known formulations. Our implementation exploits the TUM property to allow us to solve large scale phase unwrapping problems using LP solvers. Tests using simulated and real data sets demonstrate the validity of our new formulation.

In the absence of specialized constraints, illustrated in the case study above, the edgelist formulation is computationally very inefficient compared to the original MCF formulation and requires more computer resources. The quality of the solution however, matches that of the original MCF formulation and permits a simplified representation of multi-dimensional phase unwrapping problem and the incorporation of data from external sources, like GPS, where available to better constrain the unwrapped solution.

The edgelist method provides a way to apply a priori knowledge to improve our ability to unwrap challenging data sets such as our case study above. Moreover, the algorithm is easier to implement at a smaller scale with fewer number of constraints and variables. This makes it an ideal candidate for improving the temporal unwrapping step of the multi-temporal PS-InSAR and SBAS phase unwrapping [10].

Acknowledgements

We would like to thank Isabelle Ryder and Roland Burgmann from University of California, Berkeley for providing us with the initial data for PS-InSAR analysis and for useful discussions on the phase unwrapping model. This work was supported by the NASA Earth and Space Science Fellowship (NESSF).

References

11. A. P. Shanker and H. A. Zebker, “Edgelist phase unwrapping algorithm for time series InSAR analysis”, submitted to JOSA A.

**Figure 2.** (Top) This figure illustrates the additional constraint required to incorporate other geodetic observations in the phase unwrapping algorithm with the original MCF formulation and the path of integration associated with the constraint. (Bottom) This figure illustrates the same constraint in the edgelist formulation. In this case a new edge (shown in red) is introduced in the unwrapping grid (G). The TUM property of the constraint matrix is conserved in case of the edgelist formulation only.

**Figure 3.** (Left) This figure shows the skeletal framework on which the PS-InSAR data set is unwrapped. Each slice represents an interferogram in the PS-InSAR time-series. All the node potentials in areas marked by red are set to zero. The plane of the fault trace is drawn in blue. (Right) The cost of the edges cutting across the fault is subsidized as a distance of center of edge from the fault as shown.
Figure 4. (Left) Average LOS displacement rate image estimated using the edgelist phase unwrapping algorithm. The fault trace and the area used for computing the average profile is also indicated. (Right) The average LOS displacement rate (red dots) as a function of distance from the fault is shown. Assuming all the displacement was purely due to strike slip motion across the fault, we estimate a slip rate of 22 mm/yr.