Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE

Corso di Laurea Magistrale in Matematica

MATHEMATICAL MODELING: NUMERICAL SIMULATION OF LANDSLIDES

Relatore:
Chiar.ma Prof.ssa
FABIANA ZAMA

Presentata da:
LIWEI HU

Correlatore:
Chiar.ma Prof.ssa
GERMANA LANDI

Correlatore:
Chiar.mo Prof.
DAVIDE DONATI

II Sessione
Anno Accademico 2022/2023

Ringraziamenti

Ringrazio la prof.ssa Fabiana Zama e la prof.ssa Germana Landi per i preziosi consigli e per il loro supporto alla stesura di questa tesi. Un sincero ringraziamento al prof. Davide Donati per la grande disponibilità e il supporto alla stesura della parte di ingegneria geologica del lavoro.

Sommario

I movimenti franosi rappresentano un pericolo diffuso negli ambienti terrestri con pendenze, causando vittime umane nelle aree urbane, lungo le vie di trasporto e presso siti di industria rurale. Il rischio associato ai movimenti franosi e all'instabilità delle pendenze è aumentato negli ultimi decenni. I movimenti franosi innescati dall'attività umana stanno crescendo a causa della costruzione, dell'estrazione mineraria illegale e del taglio delle colline [31], oltre all'aumento della frequenza di eventi meteorologici estremi [32]. Gli eventi recenti nell'Appennino Emiliano-Romagnolo sono solo l'ennesima testimonianza dell'incremento dei movimenti franosi e dell'instabilità delle pendenze nella zona. Questo aumento del rischio ha richiamato l'attenzione sull'importanza di indagini geologiche complete per prevedere l'evoluzione delle pendenze instabili, identificare i meccanismi potenziali di cedimento e quantificare il pericolo e il rischio associati all'instabilità [1].

Stimare con precisione la velocità, la profondità, il volume e i parametri di resistenza o reologici dei movimenti franosi è essenziale per la previsione del pericolo [33]. Tuttavia, queste stime sono notoriamente difficili perché i dati superficiali e sottosuperficiali sono solitamente limitati a misurazioni puntuali e brevi finestre di tempo.

La maggior parte dei metodi classici che consentono di caratterizzare la superficie di cedimento dei movimenti franosi si basa su relazioni geometriche relativamente semplici [34]. Sebbene i metodi puramente geometrici siano facili da utilizzare, essi richiedono tipicamente forti ipotesi sulla struttura e sul comportamento del movimento franoso, e l'analisi viene effettuata solo

lungo una particolare sezione trasversale del movimento franoso, come nel metodo dell'inclinazione vettoriale [29].

L'obiettivo di questo lavoro è determinare lo spessore di una frana in roccia a partire dalle misure della velocità superficiale. In particolare, si analizza l'approccio proposto da Booth et al. (2013) in [6], che coinvolge la discretizzazione dell'equazione di trasporto su dominio bidimensionale mediante differenze finite e la risoluzione di un problema inverso per la ricostruzione dello spessore basandosi sulle misure di spostamento superficiale acquisite con la tecnica del radar ad apertura sintetica [1].

Inizialmente, vengono analizzate le caratteristiche dei dati acquisiti. Questa fase preliminare permette di comprendere appieno le peculiarità dei dati e di utilizzarli in maniera ottimale. Successivamente, viene esaminato l'approccio di regolarizzazione dal punto di vista dell'ottimizzazione numerica. Attraverso un'analisi accurata, si valutano i vantaggi e le limitazioni di tale approccio, consentendo di selezionare le soluzioni più efficaci per la ricostruzione dello spessore della frana.

Un ulteriore contributo consiste nell'approfondimento del Principio del Bilanciamento come metodo per il calcolo del parametro di regolarizzazione. Questo principio offre una base solida per determinare il parametro ottimale, permettendo di ottenere risultati accurati e affidabili nella ricostruzione dello spessore della frana.

Passando alla sintesi dei risultati ottenuti, si evidenzia un notevole miglioramento dell'efficienza computazionale rispetto all'implementazione proposta in [6]. Grazie all'ottimizzazione numerica e all'utilizzo del Principio del Bilanciamento, si riesce a ottenere una significativa riduzione dei tempi di calcolo, rendendo il processo di ricostruzione dello spessore più rapido ed efficiente. Inoltre, i risultati ottenuti sono qualitativamente in linea con le analisi effettuate tramite tecniche alternative. Ciò conferma l'affidabilità e l'accuratezza del metodo proposto, fornendo una conferma valida dell'efficacia delle tecniche di regolarizzazione utilizzate.

Infine, è stato sviluppato un'applicazione Matlab dedicata alla ricostruzione delle mappe dello spessore della frana a partire dai dati forniti. Questo strumento permette di semplificare e automatizzare il processo, consentendo agli utenti di ottenere in modo rapido e affidabile le informazioni desiderate.

La struttura della tesi è la seguente. Nel primo capitolo è introdotto il problema computazione da un punto di vista geologico e sono brevemente accennati i dati usati per l'inversione. Il capitolo 2 parte da alcuni concetti basilari di meccanica dei continui per arrivare ad una equazione che lega i dati disponibili e la stima in questione. Il capitolo 3, la parte centrale di tutto il lavoro, fornisce prima la presentazione teorica poi l'implementazione dei vari metodi numerici, usati sia per la discretizzazione dell'equazione differenziale, sia per la regolarizzazione del problema con i moderni metodi di ottimizzazione. Nel capitolo 4 sono illustrati gli esperimenti numerici condotti con i vari paramatri selezionati. Nel capitolo 5 sono richiamate le ipotesi che i geologi hanno formulato sulla struttura della frana a partire da studi sul campo, analisi di profili, confronto con dati con l'altimetria laser [1] e applicazione di l'analisi delle componenti indipendenti [30]. Infine nel capitolo conclusivo è data una possibile interpretazione dei risultati numerici in relazione alle conoscenze attuali sulla struttura della frana e ulteriori possibili sviluppi del metodo.

Abstract

Landslides are an ubiquitous hazard in terrestrial environments with slopes, along transport corridors and at sites of rural industry. The risk associated with landslides has increased over the past decades, both in relation to human activity [31] and to climate change [32]. The recent events in the Emilia-Romagna Apennines are just the latest testament to the increasing occurrence of landslides and slope in the area. This increased risk has called attention to the importance of comprehensive geological investigations to forecast the evolution of unstable slopes, identify potential failure mechanisms, and quantify the hazard and risk associated with the instability.

Accurately estimating landslides' velocity, depth, volume, and strength or rheological parameters is essential for hazard prediction [33]. However, these estimates are notoriously difficult because surface and subsurface data are usually limited to point measurements.

Most of the classical methods that allow us to characterize the landslide failure surface geometry are based on relatively simple geometric relations [34]. Although purely geometrical methods are easy to use, they typically require strong assumptions on the structure and behavior of the landslide, and the analysis is carried out only along a particular cross-section of the slope, as in the vector inclination method [29].

In this work, we determine the thickness of a rock compound slide in Alaska [1] from measurements of superficial velocity. In particular, we analyze the method proposed by Booth et al. (2013) in [6], which involves the discretiza-

tion of a transport equation with bidimensional domain by means of finite difference approximations and the solution of an inverse problem to infer the underlying thickness from superficial displacements acquired using a synthetic aperture radar [1].

We first analyze the features of the acquired data. This preliminary phase allows us to comprehend the peculiarity of data and to understand how to use them properly. Then, we study the regularization techniques and the optimization algorithms used to infer the thickness. An accurate analysis shows the advantages and limitations of these methods, allowing to identify the most efficient and reliable implementation.

A further contribution consists in the investigation of the Balancing Principle as the criterion to select the regularization parameter. This principle provides a solid foundation to determine the proper regularization parameter, allowing us to obtain accurate and realistic results in the estimation of depth of Fels landslide.

Regarding the results, a remarkable improvement in efficiency has been achieved compared to the method proposed in [6]. Thanks to more efficient optimization algorithms and the application of the Balancing Principle, it's possible to reduce drastically the computation time making the process of inversion more practical. The results obtained with Fels landslide data are qualitatively consistent with other analysis techniques. Thus the reliability and the accuracy of the proposed method are confirmed, proving also the effectiveness of the applied regularization techniques.

Finally, we developed a Matlab application to synthesize the depth of a landslide from input data. This tool allows to simplify and automate the inversion process, providing users with the estimate rapidly and effectively.

The thesis is structured as follows. In Chapter 1, we introduce the computational problem from a geological point of view and we shortly present the data used for inversion. Chapter 2 provides some basic notions of continuum

mechanics which allow us to derive a relation between data and the unknown depth. Chapter 3 forms the central part of the work, we first present the theoretical aspects of numerical methods to approximate differential equations, notions of inverse theory and modern optimization algorithms, then we discuss their implementations. In Chapter 4, results obtained with numerical experiments and Fels landslide data are shown. In Chapter 5, we review the latest results engineering geologists have found on the structure of this slope from field studies, profile analysis, comparison with data acquired from airborne laser scanning [1] and application of independent component analysis [30]. Finally, in Chapter 6 we give a possible interpretation of numerical results in relation to the hypotheses discussed in the previous chapter, and further possible developments to improve the method.

Contents

Chapter 1

The Geological Problem

1.1 Problem Definition

The risk associated with landslides and slope instability has increased over the past several decades due to expanded development of mountainous areas to accommodate population and tourism growth as well as increased frequency of extreme weather events resulting from ongoing climate change [1]. This increased risk has called attention to the importance of comprehensive geological investigations to forecast the evolution of unstable slopes.

The hazardous impact and erosive potential of slow moving landslide depends on landslide properties including velocity, size, and frequency of occurrence. Importantly, the landslide failure style also impacts our ability to measure landslide properties, such as thickness and volume, which can strongly influence runout and erosion rate [2]. Some landslides create clear and identifiable scars and deposits by evacuating material from hillslope, making it possible to directly measure landslide properties from field data, digital elevation models and remote sensing observations. However, for landslides that move slowly for years or centuries, referred to as **slow-moving landslides**, and do not create hillslope scars, it is difficult to infer their thickness and volume.

Direct measurements can be taken from boreholes, but data are usually lim-

ited to isolated point measurements, for which financially expensive and time consuming operations are required. As a consequence, the collected data cannot capture the spatial variability exhibited by these landslides.

An example of these landslide is the Fels landslide. **Fels landslide** is a rock-compound landslide located in the Alaska Range about 150 km southeast of Fairbanks, on the north slope of Fels Glacier valley, which is tributary of the Delta River valley (Figure 1.1).

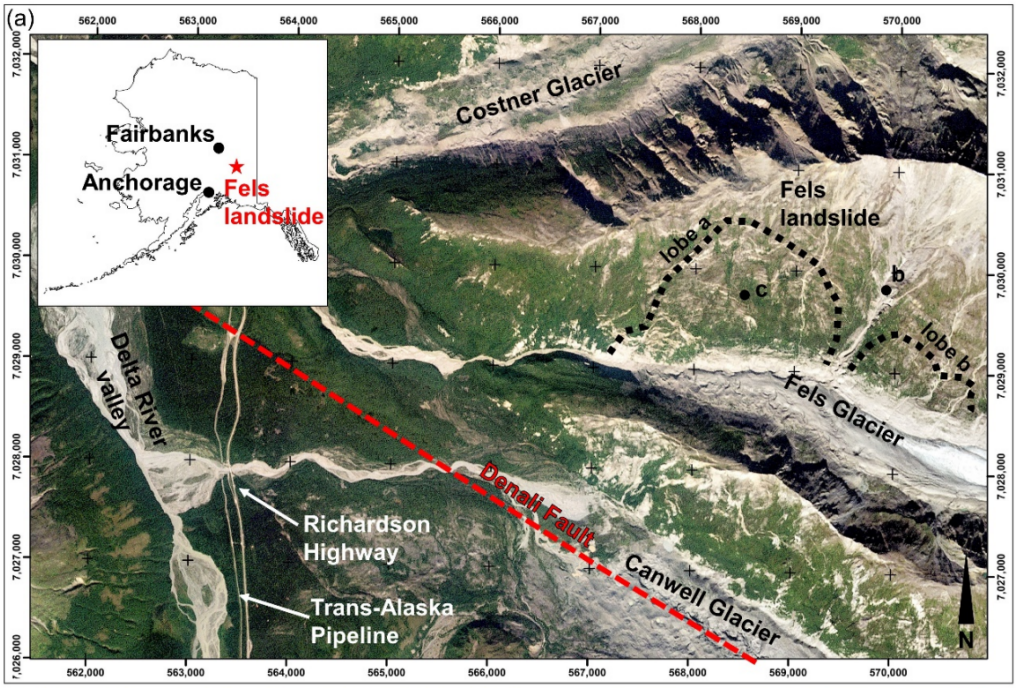


Figure 1.1: 2017 satellite view of the intersection of the Delta River valley and Fels valley. The black-dotted lines delineate the areas of ongoing slope deformation within Fels Glacier valley. The red dashed line shows the trace of the Denali Fault. RapidEye 4, WGS84/UTM 6N grid. Picture from [1].

It terminates at an elevation of 820 m above sea level.
The glacier has been retreating for more than a century and a comparison of aerial photographs taken in 1949 and 2017 shows that the glacier retreated more than 900 m over this period (Figure 1.2).

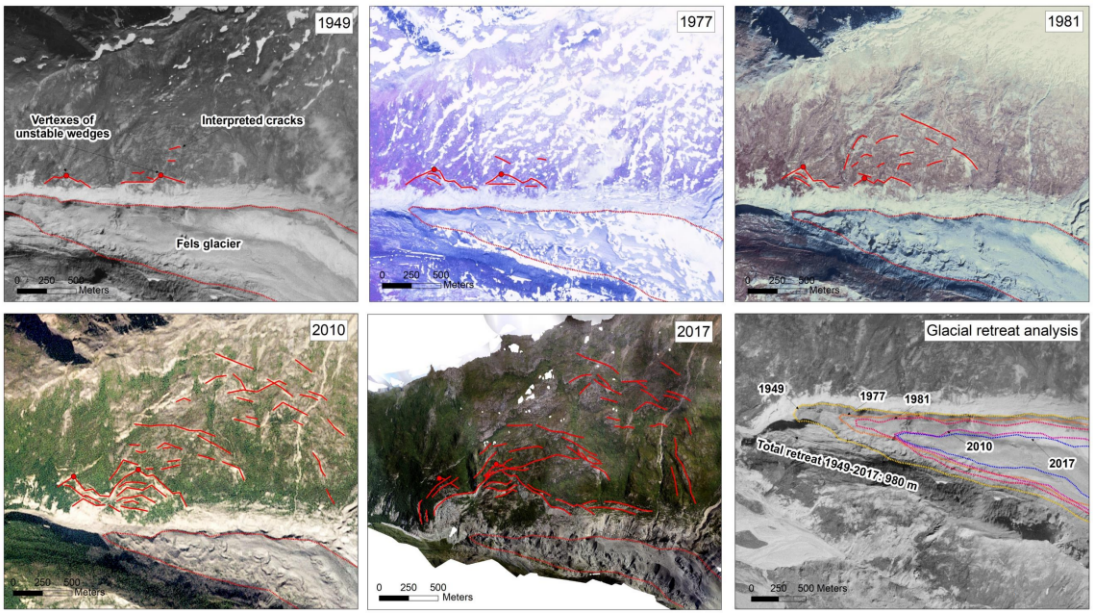


Figure 1.2: Aerial photographs of Fels Glacier from 1949 to 2017. Courtesy of Dr. Davide Donati.

Within Fels Glacier valley, slope deformation involves 2 areas, referred to as **lobe a** and **lobe b** (Figure 1.1) , that are divided by a deeply incised gully, perpendicular to the valley. Lobe a displays significantly higher deformation rates and prominent slope damage features, compared to lobe b. In the following, we consider lobe a as Fels landslide.

Fels landslide has a surface area of about 2.3 km^2 and extends 1400 m in the East-West direction and 1600 m in the North-South direction between 920 and 1490 m above sea level. The ground surface within the landslide area has an overall slope of 20-30° to the South, steepening up to a slope angle of 40-50° at its glacially eroded toe. The rock mass has a prominent slope-parallel foliation (D1 in figure 1.3 (b)), and two other orthogonal discontinuity sets (D2 and D3 in figure 1.3 (b)). Much of the slope surface is covered by colluvium that hosts localized superficial slumps (figure 1.3 (c)).

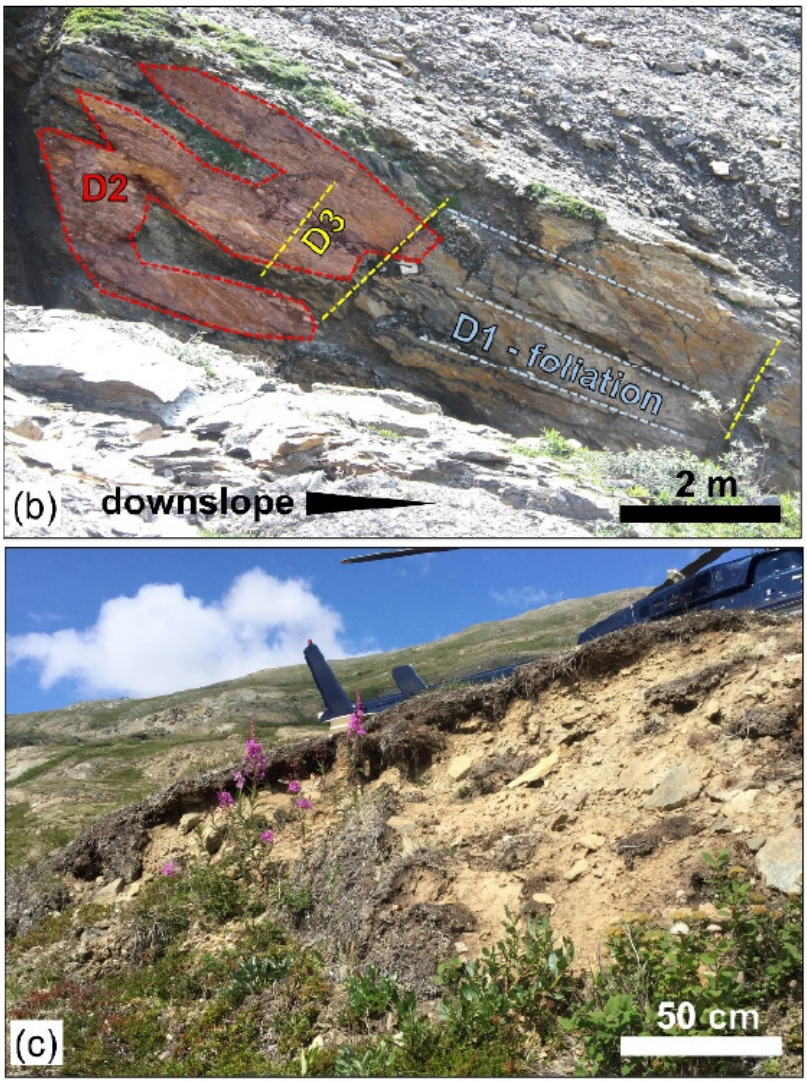


Figure 1.3: (b) Detail of the rock mass forming the slope. Note the slope-parallel foliation D1 and the discontinuity sets: D2, which forms the outcrop surface (red, dashed outline shows a D2 discontinuity plane), and D3, dipping into the slope. (c) View of a surface slump scar, displaying the colluvial material drapping the slope. Pictures from [1].

Glacial deposits are associated with lateral and frontal moraines in the lower part of Fels valley and in the Delta River valley. The active Denali Fault is located 3 km south of the Fels landslide (Figure 1.1). It strikes

NW-SE and follows and controls the orientation of the Canwell valley, which neighbors Fels valley on the south.
An earthquake of moment magnitude 7.9 occurred on the Denali Fault on November 3, 2002. Many co-seismic rock avanlanches, rockslides, slumps, and debris avanlanches were triggered by the earthquake in the epicentral area. Although no major landslides occurred within Fels valley, the intense ground shaking, exacerbated by material and topographic amplification, is likely to have enhanced internal damage (e.g. rock mass dilatation, fracture propagation) within the slope.
Historical, geological, and geomechanical analyses indicate that the current instability phase in the Fels Glacier valley was initiated by glacier retreat, which provided kinematic freedom for movement of the toe of the landslide. Slumping at the toe, in turn, caused the instability to propagate upslope, where displacements are occurring by planar sliding along a slope parallel rupture surface that is likely controlled by foliation.

1.2 Available Data

Direct measurements of landslide thickness are expensive and could be available only at a limited number of points. An alternative way to study the failure style of a landslide is provided by computer simulations. Typically, a computer simulation is based on a physical model which has relevant physical quantities constrained by data derived from fieldwork or remote sensing surveys.

A wealth of satellite sensors can be used to study landslides [1]. Optical sensors provide multispectral and panchromatic imagery as well as hyperspectral datasets.
Synthetic aperture radar (SAR) processing algorithms create images from microwave signals emitted and received after scattering from Earth's surface, by sensors mounted on satellite-based, aerial-based or ground-based platforms. Both the amplitude and phase of the radar waves backscattered

along the line of sight (LoS) are recorded by these sensors.
Millimeter-scale displacements of the surface can be measured by computing the interferometric phase difference of multiple such SAR images with the same line of sight geometry acquired at different time (Figure 1.4). This technique is known as **Differential Interferometric SAR** (DInSAR).

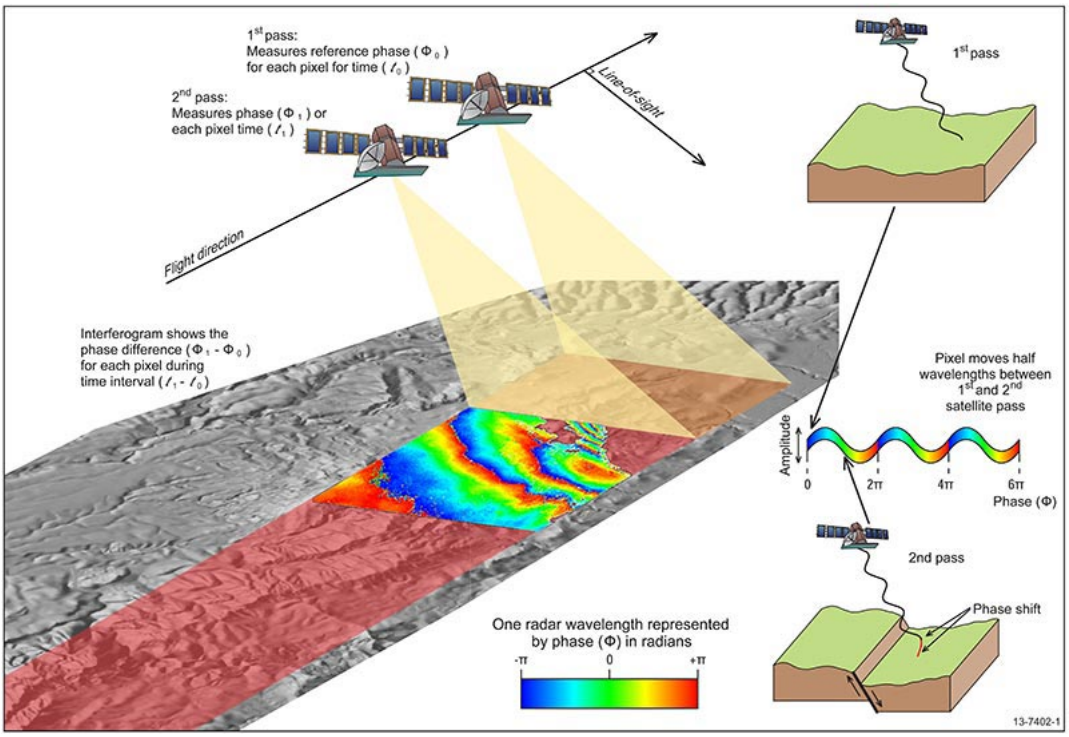


Figure 1.4: An interferogram created by using two SAR images that are acquired in two different times and maps the phase shift caused by a movement of the surface between the two acquisitions - Picture from https://www.un-spider.org/links-and-resources/data-sources/daotm-land-deformation.

When data from different viewing geometries are analyzed, the 3D decomposition of the DInSAR measurements can provide the magnitude of the displacements in the vertical and horizontal directions, by solving basically a least-square problem [19]. In all cases, these approaches involve analysis of datasets acquired along at least 2 distinct line of sight geometries by a

satellite moving along **ascending paths** (satellite traveling approximately from south S to north N, looking east E) and **descending paths** (satellite traveling approximately from N to S, looking west W).

An important limitation of the DInSAR technique is its relative insensitivity to the N-S component of slope deformation, due to the E-W LoS of space-borne SAR sensors, which are constrained to **polar orbits** (Figure 1.5).

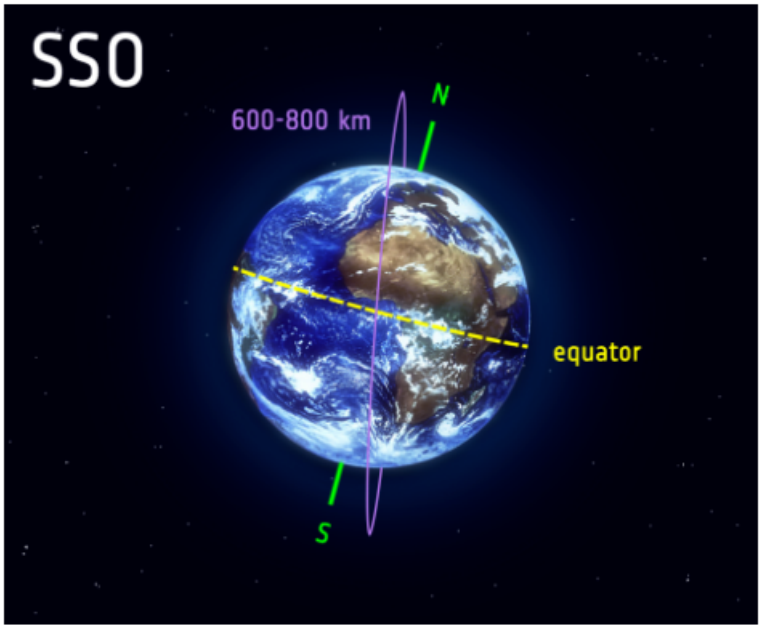


Figure 1.5: **Sun-synchronous orbit** (SSO), a particular kind of polar orbit - Picture from https://www.esa.int/Enabling_Support/Space_Transportation/Types_of_orbits.

Therefore the true 3D displacement vector can only be determined using sophisticated analyses of several LoS geometries, or estimated based on constraints imposed by slope orientation.
The error statistics of the multi-DInSAR inversions of 3D displacement are affected by many factors. The most important of these factors include temporal coherence losses of the individual interferograms, imperfect mutual overlap of the temporal intervals of the individual interferograms (if displacement

changes over time), as well as the robustness of the multiple LoS geometries spanning 3D spaces. For the case of solely using multiple LoS without additional constraints (such as surface parallel displacement) the geometric robustness, expressed by the determinant of the least square inversion matrix, leads to poor error statistics. This as the available diversity of LoS for spaceborne sensors is limited leading to multi-LoS geometries with small inversion determinants far from ideal case of 3 orthogonal LoS.
Nevertheless, DInSAR is a powerful tool for studying landslides and is routinely employed to monitor slope displacements associated with slow and very slow landslides.

A further limitation of conventional DInSAR monitoring (using a single pair of images) is that, to maintain signal quality and avoid temporal decorrelation in non-arid areas the technique can only be used to monitor over relatively short periods of days or weeks. To overcome this limitation, various approaches have been proposed. Digital image correlation (DIC) allows tracking of pixel blocks in co-registered 2D images to map and quantify changes between the images, allowing more rapid movements to be monitored. The DIC technique has been applied to track changes in optical satellite imagery and in SAR scenes, using offset-tracking technique methods.

The SAR speckle tracking (ST) technique is a particular application of DIC methods. SAR ST algorithms exploit the amplitude of the SAR scenes and extend the ability of conventional DInSAR to provide deformation measurements of up to tens of meters, depending on the resolution of the data. For a given pixel size of a SAR image pair the spatial resolution of the resulting ST map is considerably less (by a factor of 50 or more) than that of a DInSAR map. This requires higher resolution of the primary SAR data to capture reasonable spatial detail of slide motion with the ST method.

In the case of Fels landslide, Donati et al. (2021) presented in [1] a procedure that exploits the deformation field extracted by ST algorithms from multi-geometry datasets to characterize the progressive (substantial) defor-

mation that occurs over a 10-year period. They derived a spatial dataset comprising full 3D displacement vectors using very high (sub-meter) resolution RADARSAT-2 (Figure 1.6) Spotlight mode scene acquired from both ascending and descending orbits.

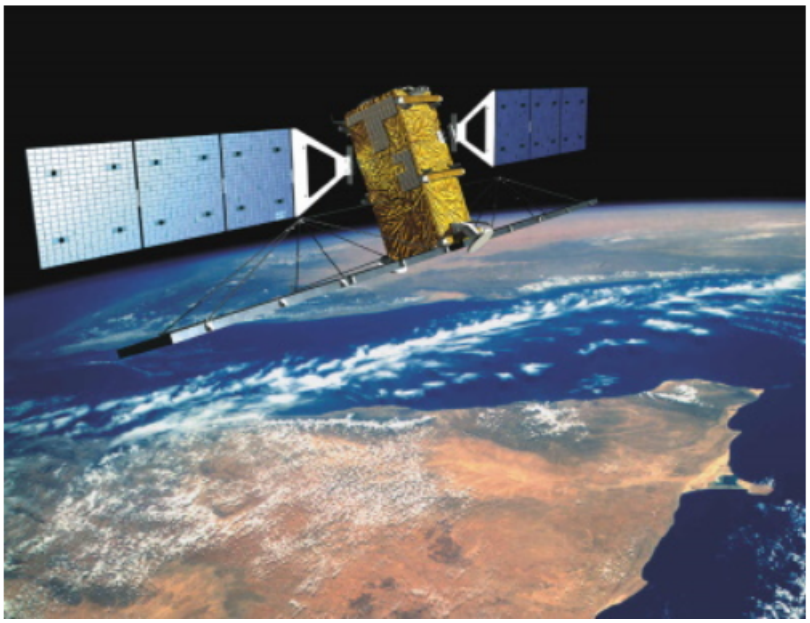


Figure 1.6: RADARSAT-2: a Canadian Space Agency Earth observation satellite - Picture from
https://earth.esa.int/eogateway/missions/radarsat.

The analysis was focused on progressive slope deformation over 2 subsequent 5-year periods (2010-2015 and 2015-2020).

Acquisition Date	Beam Mode	Orbit Direction	Incidence Angle
14 July 2010	Spotlight-U19	Descending	44.0°
12 July 2015	Spotlight-U19	Descending	44.0°
2 August 2020	Spotlight-U19	Descending	44.0°
15 July 2010	Spotlight-U20	Ascending	44.6°
13 July 2015	Spotlight-U20	Ascending	44.6°
3 August 2020	Spotlight-U20	Ascending	44.6°

Figure 1.7: Table showing RADARSAT-2 data acquisition parameters - Picture from [1].

Chapter 2

Physical Modeling

2.1 The Continuous Model

From the superficial displacement field, we can obtain an approximation of the real mean annual velocity field. Surprisingly from these data alone, we'll show in this chapter and the next one that it is possible to give an estimate of the landslide thickness, therefore of the failure plane geometry. First of all, we need to find a relation between the available data and the thickness. Since a generic landslide can be approximated, for simplicity, by a continuous body, in the following some basic definitions from Continuum Mechanics [3] will be recalled.

Definition 1. *A* ***continuous body*** *B is a regular open set of the ordinary space $\mathbb{R}^3$. Each point of B, denoted by P, identifies a material point of the body and similarly each subset A, a part of the body.*

Properly speaking this open set is the initial configuration of a physical body with respect to an inertial Cartesian reference frame F,

$$B \subseteq \{t_0\} \times \mathbb{R}^3$$

so it's more appropriate to call it **reference configuration** of the body. But from now on when we say continuous body, we mean its initial configuration

with the related frame.

Fels Landslide can be considered as a continuous body. We'll set the initial time as July 12, 2015 the date the first measurements of displacement are taken in the 2015-2020 time window. We will also consider an inertial Cartesian frame such that the x-direction corresponds to E-W direction (with E positive), the y-direction corresponds to N-S direction (with N positive) and the z-direction corresponds to the vertical direction (with upward direction positive). Finally, for visualization purposes, it's convenient to assume the point cloud of superficial displacement lies within the positive orthant, in particular we'll see that the problem can be simplified as a bivariate one such that the discrete grid lies within the positive quadrant.

Definition 2. *Given a continuous body B, a **deformation** f on B is any transformation*

$$f : B \to \mathbb{R}^3$$

such that

(i) f is smooth

$$f \in C^\infty$$

(ii) f is 1-1

(iii) f has positive-determinant Jacobian

$$\forall P \in B \quad detJ_f(P) > 0$$

Definition 3. *Given a continuous body B, a **motion** $\vec{x}$ is a 1-parameter family of deformations*

$$\vec{x} : (0, \infty) \times B \to \mathbb{R}^3$$

which is smooth

$$\vec{x} \in C^\infty$$

Since for each material point P at instant t, the vector $\vec{x}(t, P)$ represents its spatial position at time t (Figure 2.1), the set of all $(t,\vec{x})$ is called ***trajectory*** *T of the motion*

$$T = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 | (x_1, x_2, x_3, x_4) = (t, \vec{x}(t, P)), t \in (0, \infty) \wedge P \in B\}$$

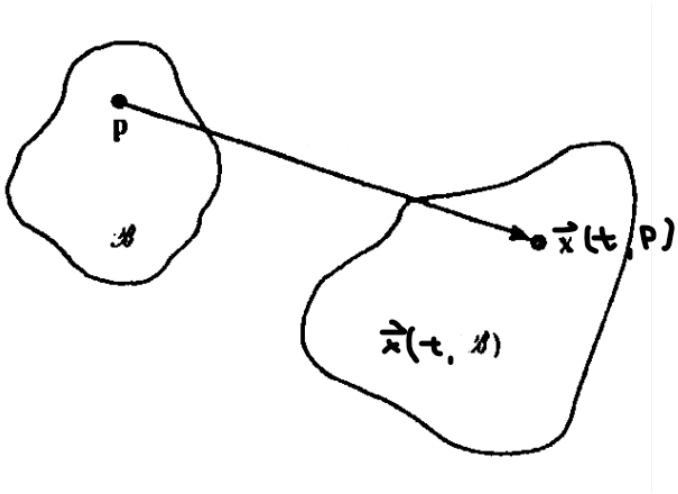


Figure 2.1: A material point P of a continuous body and its position at time t.

The **velocity** of the body associated to the motion is its time derivative.

$$\vec{v} = \frac{\partial \vec{x}}{\partial t}$$

similarly the **acceleration**, its second time derivative

$$\vec{a} = \frac{\partial^2 \vec{x}}{\partial t^2}.$$

The natural motion of the Fels landslide initiated by glacier retreat at the toe is indeed smooth. We will assume that each **instant deformation** $\vec{x}(t, \cdot)$ is injective, having also positive-determinant Jacobian.

Definition 4. *A continuous body B has* ***continuously distributed mass*** *if for any deformation f there's a density function ρ_f*

$$\rho_f : Im(f) \to \mathbb{R}$$

such that

(i) ρ_f *is smooth*

(ii) ρ_f *is positive*

(iii) the integral of ρ_f *over any deformed part of the body is the mass of that part*

$$\forall A \subseteq B \quad \int_{f(A)} \rho_f = m(A)$$

We will assume the Fels Landslide has continuously distributed mass. The main result from Continuum Mechanics that we'll use is the following.

Lemma 1 (Local Conservation of Mass)**.** *Given a continuous body* B *in motion* $\vec{x}$*, if* B *has continuously distributed mass, by setting* ρ *as* ***density in the motion***

$$\rho : T \to \mathbb{R}$$
$$\rho(t, \vec{x}) = \rho_{\vec{x}(t,.)}(\vec{x}),$$

where $\rho_{\vec{x}(t,.)}$ *is the density function associated with* $\vec{x}(t, .)$*, the deformation at time* t*, and by introducing the* ***reference map*** P *as*

$$P : T \to B$$
$$P(t, \vec{x}) = (\vec{x}(t, .))^{-1}(\vec{x})$$

then it can be proved that the following relation between the density and the velocity must hold during motion

$$\forall (t, \vec{x}) \in T \quad \frac{\partial \rho}{\partial t}(t, \vec{x}) + \rho(t, \vec{x}) < \vec{\nabla}_{\vec{x}}, \vec{v}(t, P(t, \vec{x})) >= 0$$

Lemma 2. *In particular, if the motion is* ***isochoric*** *i.e.*

$$\frac{\partial \rho}{\partial t} \equiv 0$$

the relation becomes

$$\forall (t, \vec{x}) \in T \quad < \vec{\nabla}_{\vec{x}}, \vec{v}(t, P(t, \vec{x})) >= 0$$

2.2 Thickness Equation

A glacier, like many other entities, can also be approximated by a continuous body with continuously distributed mass. The following result is known in Ice Dynamics [4] [20] as the ice-thickness equation.

Definition 5. *The surface of a glacier, seen as a continuous body B with its frame, whose velocity field can actually be measured, is known as the* ***free surface*** *and in the inertial Cartesian reference frame, at each time t, it can be represented in implicit form as*

$$\{(t,x,y,z) \in \{t\} \times \mathbb{R}^3 : F_s(t,x,y,z) = 0\}$$

where

$$F_s(t,x,y,z) = z - f(t,x,y).$$

Similarly the ***base surface*** *has implicit representation at a particular time t*

$$\{(t,x,y,z) \in \{t\} \times \mathbb{R}^3 : F_b(t,x,y,z) = 0\}$$

where

$$F_b(t,x,y,z) = z - b(t,x,y).$$

Consequently, we can define the ***underlying thickness*** *h at time t as*

$$\begin{aligned} &h : Proj^t \to \mathbb{R} \\ &h(t,x,y) = f(t,x,y) - b(t,x,y) \end{aligned}$$

where $Proj^t \subseteq \{t\} \times \mathbb{R}^3$ is the ***projection of the free surface*** *on $\{z = 0\}$ at time t.*

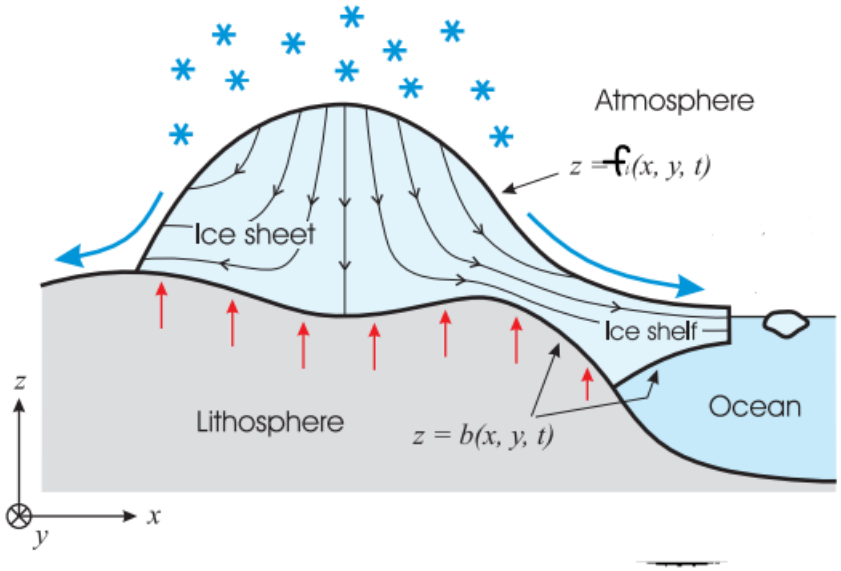


Figure 2.2: The thickness of a glacier as difference of 2 heights - Picture from [4].

From this definition it follows that the thickness is independent from the height variable z. It's not difficult to see that the previous definitions are well-defined for a generic continuous body, in particular for a landslide, under the assumption that the configuration is similar to the one in Figure 2.2.

The goal of following work is to give an estimate of the thickness of the Fels landslide on $Proj^{t_0}$ where t_0=July 12, 2015.

Theorem 1 (Ice-Thickness Equation)**.** *Given an active glacier with continuously distributed mass in isochoric motion, by setting, at each time t, the* ***ice-flux*** *Q as*

$$Q : Proj^t \to \mathbb{R}^2$$

$$(Q_x, Q_y)(t,x,y) = \left(\int_{b(t,x,y)}^{f(t,x,y)} v_x(t, P(t,x,y,z))dz, \int_{b(t,x,y)}^{f(t,x,y)} v_y(t, P(t,x,y,z))dz\right),$$

then the ice-flux divergence is balanced by the rate of thickness change and the net surface/basal mass balances, namely

$$\forall(t,x,y) \in Proj^t$$

$$\begin{aligned}\frac{\partial h}{\partial t}(t,x,y) =& - < \nabla_{(x,y)}, Q(t,x,y) > \\ & - \frac{\partial}{\partial t}(F_s(t,x,y,P(t,x,y,f(t,x,y)))) \\ & + \frac{\partial}{\partial t}(F_b(t,x,y,P(t,x,y,b(t,x,y))))\end{aligned}$$

where f(t,x,y) is the height of the free surface and b(t,x,y) is the height of the basal surface.

Proof. From Lemma 2 we know that

$$\forall(t,x,y,z) \in T$$

$$\frac{\partial}{\partial x}(v_x(t,P(t,x,y,z))) + \frac{\partial}{\partial y}(v_y(t,P(t,x,y,z))) + \frac{\partial}{\partial z}(v_z(t,P(t,x,y,z))) = 0$$

in particular, at each time t, the vertical integral from the basal height b to the surface height f of the same function is zero

$$\forall(t,x,y) \in Proj^t$$

$$\int_{b(t,x,y)}^{f(t,x,y)} \frac{\partial}{\partial x}(v_x(t,P(t,x,y,z))) + \frac{\partial}{\partial y}(v_y(t,P(t,x,y,z))) + \frac{\partial}{\partial z}(v_z(t,P(t,x,y,z)))dz = 0$$

The third integral, according to the fundamental theorem of calculus, is simply

$$\int_{b(t,x,y)}^{f(t,x,y)} \frac{\partial v_z}{\partial z}(t,P(t,x,y,z))dz = v_z(t,P(t,x,y,f(t,x,y))) - v_z(t,P(t,x,y,b(t,x,y))) \tag{2.1}$$

For Leibniz integral rule, we have

$$\begin{aligned}\frac{\partial}{\partial x}(\int_{b(t,x,y)}^{f(t,x,y)} v_x(t,P(t,x,y,z))dz) =& \frac{\partial f}{\partial x}(t,x,y)v_x(t,P(t,x,y,f(t,x,y))) \\ & - \frac{\partial b}{\partial x}(t,x,y)v_x(t,P(t,x,y,b(t,x,y))) \\ & + \int_{b(t,x,y)}^{f(t,x,y)} \frac{\partial}{\partial x}(v_x(t,P(t,x,y,z)))dz\end{aligned}$$

$$\frac{\partial}{\partial y}(\int_{b(t,x,y)}^{f(t,x,y)} v_y(t,P(t,x,y,z))dz) = \frac{\partial f}{\partial y}(t,x,y)v_y(t,P(t,x,y,f(t,x,y))) - \frac{\partial b}{\partial y}(t,x,y)v_y(t,P(t,x,y,b(t,x,y))) + \int_{b(t,x,y)}^{f(t,x,y)} \frac{\partial}{\partial y}(v_y(t,P(t,x,y,z)))dz$$

The first and second integrals in the mass conservation equation, being the last addend in the right hand side of each equation, can be written respectively as

$$\int_{b(t,x,y)}^{f(t,x,y)} \frac{\partial}{\partial x}(v_x(t,P(t,x,y,z)))dz = -\frac{\partial f}{\partial x}(t,x,y)v_x(t,P(t,x,y,f(t,x,y))) \frac{\partial b}{\partial x}(t,x,y)v_x(t,P(t,x,y,b(t,x,y))) \frac{\partial}{\partial x}(\int_{b(t,x,y)}^{f(t,x,y)} v_x(t,P(t,x,y,z))dz) \tag{2.2}$$

$$\int_{b(t,x,y)}^{f(t,x,y)} \frac{\partial}{\partial y}(v_y(t,P(t,x,y,z)))dz = -\frac{\partial f}{\partial y}(t,x,y)v_y(t,P(t,x,y,f(t,x,y))) \frac{\partial b}{\partial y}(t,x,y)v_y(t,P(t,x,y,b(t,x,y))) \frac{\partial}{\partial y}(\int_{b(t,x,y)}^{f(t,x,y)} v_y(t,P(t,x,y,z))dz). \tag{2.3}$$

Substituting the expressions of the three integrals, namely (2.2), (2.3) and (2.1) , in the local conservation of mass equation and by recalling the definition of ice-flux Q, we get the following equation

$$\begin{aligned} 0 = &< \nabla_{(x,y)}, Q(t,x,y) > \\ &- \frac{\partial f}{\partial x}(t,x,y)v_x(t,P(t,x,y,f(t,x,y))) - \frac{\partial f}{\partial y}(t,x,y)v_y(t,P(t,x,y,f(t,x,y))) \\ &+ v_z(t,P(t,x,y,f(t,x,y))) \\ &+ \frac{\partial b}{\partial x}(t,x,y)v_x(t,P(t,x,y,b(t,x,y))) + \frac{\partial b}{\partial y}(t,x,y)v_y(t,P(t,x,y,b(t,x,y))) \\ &- v_z(t,P(t,x,y,b(t,x,y))) \end{aligned}$$

On the other hand, by definition of surface height and basal height

$$F_s(t,x,y,z) = z - f(t,x,y)$$

$$F_b(t,x,y,z) = z - b(t,x,y)$$

writing the explicit dependence on material points

$$F_s(t,x(t,P),y(t,P),z(t,P)) = z(t,P) - f(t,x(t,P),y(t,P))$$

$$F_b(t,x(t,P),y(t,P),z(t,P)) = z(t,P) - b(t,x(t,P),y(t,P))$$

and deriving with respect to time

$$\frac{\partial}{\partial t}(F_s(t,x(t,P),y(t,P),z(t,P))) = \frac{\partial}{\partial t}z(t,P) - \frac{\partial}{\partial t}f(t,x(t,P),y(t,P))$$

$$\frac{\partial}{\partial t}(F_b(t,x(t,P),y(t,P),z(t,P))) = \frac{\partial}{\partial t}z(t,P) - \frac{\partial}{\partial t}b(t,x(t,P),y(t,P))$$

Leibniz rule yields

$$\begin{aligned}\frac{\partial}{\partial t}(F_s(t,x(t,P),y(t,P),z(t,P))) = v_z(t,P) - (&\frac{\partial f}{\partial t}(t,x(t,P),y(t,P)) \\ &+ \frac{\partial f}{\partial x}(t,x(t,P),y(t,P))v_x(t,P) \\ &+ \frac{\partial f}{\partial y}(t,x(t,P),y(t,P))v_y(t,P))\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial t}(F_b(t,x(t,P),y(t,P),z(t,P))) = v_z(t,P) - (&\frac{\partial b}{\partial t}(t,x(t,P),y(t,P)) \\ &+ \frac{\partial b}{\partial x}(t,x(t,P),y(t,P))v_x(t,P) \\ &+ \frac{\partial b}{\partial y}(t,x(t,P),y(t,P))v_y(t,P)).\end{aligned}$$

In particular for a material point belonging to the free surface $P(t,x,y,f(t,x,y))$, which at time t has height $f(t,x,y)$, we have

$$\begin{aligned}\frac{\partial}{\partial t}(F_s(t,x,y,f(t,x,y))) + \frac{\partial f}{\partial t}(t,x,y) =& v_z(t,P(t,x,y,f(t,x,y))) \\ &- \frac{\partial f}{\partial x}(t,x,y)v_x(t,P(t,x,y,f(t,x,y))) \\ &- \frac{\partial f}{\partial y}(t,x,y)v_y(t,P(t,x,y,f(t,x,y)))\end{aligned}$$

and similarly for a material point belonging to the base surface $P(t,x,y,b(t,x,y))$, which at time t has height $b(t,x,y)$, we have

$$\begin{aligned}\frac{\partial}{\partial t}(F_b(t,x,y,b(t,x,y))) + \frac{\partial b}{\partial t}(t,x,y) =& v_z(t,P(t,x,y,b(t,x,y))) \\ &- \frac{\partial b}{\partial x}(t,x,y)v_x(t,P(t,x,y,b(t,x,y))) \\ &- \frac{\partial b}{\partial y}(t,x,y)v_y(t,P(t,x,y,b(t,x,y))).\end{aligned}$$

Therefore the final form of mass conservation equation is

$$\begin{aligned}0 =& < \nabla_{(x,y)}, Q(t,x,y) > \\ &+ \frac{\partial}{\partial t}(F_s(t,x,y,P(t,x,y,f(t,x,y)))) + \frac{\partial f}{\partial t}(t,x,y) \\ &- \frac{\partial}{\partial t}(F_b(t,x,y,P(t,x,y,b(t,x,y)))) - \frac{\partial b}{\partial t}(t,x,y)\end{aligned}$$

or equivalently

$$\begin{aligned}\frac{\partial f}{\partial t}(t,x,y) - \frac{\partial b}{\partial t}(t,x,y) =& - < \nabla_{(x,y)}, Q(t,x,y) > \\ &- \frac{\partial}{\partial t}(F_s(t,x,y,P(t,x,y,f(t,x,y)))) \\ &+ \frac{\partial}{\partial t}(F_b(t,x,y,P(t,x,y,b(t,x,y))))\end{aligned}$$

Recall the definition of the thickness to conclude the proof. □

The last two addends of the right hand side are called respectively **surface mass balance** (precipitations might increase the height of the free surface) and **basal mass balance** (ice melting will decrease overall the thickness). For a landslide these two terms are typically negligible [6], so thickness equation can be simplified

$$\forall (t,x,y) \in Proj^t \quad \frac{\partial h}{\partial t}(t,x,y) = - < \nabla_{(x,y)}, Q(t,x,y) > .$$

Moreover, by the definition of volume flux, if we set $(v_x^{(z-mean)}, v_y^{(z-mean)})$

as the **vertically-averaged horizontal velocity**, the ice-flux Q can be approximated by

$$Q(t,x,y) \simeq (hv_x^{(z-mean)}, hv_y^{(z-mean)})(t,x,y),$$

where each integral in the definition of Q yields the thickness h. Therefore, the thickness equation can be seen as

$$\forall (t,x,y) \in Proj^t \quad \frac{\partial h}{\partial t}(t,x,y) = - < \nabla_{(x,y)}, h(v_x^{(z-mean)}, v_y^{(z-mean)}) > (t,x,y).$$

If the landslide is thin relative to its length then the vertically-averaged horizontal $(v_x^{(z-mean)}, v_y^{(z-mean)})$ should be a fraction of the **surface horizontal velocity** $(v_x^{(surf)}, v_y^{(surf)})$ [5]:

$$v_x^{(z-mean)}(t,x,y) = \gamma v_x^{(surf)}(t,x,y)$$

$$v_y^{(z-mean)}(t,x,y) = \gamma v_y^{(surf)}(t,x,y)$$

with $\gamma \in (0,1)$. The latter is exactly part of the data available and can be visualized as a vector field at the free surface of the landslide.

Finally, assuming that the basal height is constant within the time interval in which measurements are available

$$\frac{\partial b}{\partial t}(t,x,y) = 0$$

and the landslide does not behave like a rigid body, the change in thickness can be approximated by the surface vertical velocity

$$v_z^{(surf)}(t,x,y) = \frac{\partial f}{\partial t}(t,x,y).$$

By setting the **scaled thickness** h_γ as

$$h_\gamma = \gamma h,$$

the relation we were looking for is

$$\forall (t,x,y) \quad v_z^{(surf)}(t,x,y) = - < \nabla_{(x,y)}, h_\gamma(v_x^{(surf)}, v_y^{(surf)}) > (t,x,y).$$

To summarize the previous remarks we can state the following theorem analogous to Theorem 1.

Theorem 2 (Landslide Thickness Equation)**.** *Given an active landslide with continuously distributed mass in isochoric motion, which is thin relative to its length, assuming that the basal height is constant and the landslide does not behave like a rigid body, if the surface velocity field at a particular time t_0 can be measured (indirectly)*

$$\vec{v}^{(surf)} = (v_x^{(surf)}, v_y^{(surf)}, v_z^{(surf)})$$

then an estimate of the underlying thickness field h can be given by solving the following partial differential equation (PDE)

$$v_z^{(surf)} = - < \nabla_{(x,y)}, (h_\gamma v_x^{(surf)}, h_\gamma v_y^{(surf)}) > \tag{2.4}$$

where $\gamma \in (0, 1)$ is a constant number which depends on landslide rheology.

Although landslides might have heterogeneous material properties, this is rarely quantified, and defining γ as constant implies that the landslide's rheology is spatially uniform so that changes in thickness alone are responsible for the observed deformation field.

Chapter 3

Numerical Modeling

3.1 The Discretized Model

Writing explicitly the equation (2.1), i.e. applying Leibniz rule, the sought relation between data and thickness becomes

$$-v_z^{(surf)} = \frac{\partial h_\gamma}{\partial x} v_x^{(surf)} + \frac{\partial v_x^{(surf)}}{\partial x} h_\gamma + \frac{\partial h_\gamma}{\partial y} v_y^{(surf)} + \frac{\partial v_y^{(surf)}}{\partial y} h_\gamma. \tag{3.1}$$

If the superficial velocities $\vec{v}^{(surf)}$ were completely known, in order to find the scaled thickness all we need to do would be solving a first order PDE in two variables. But since the superficial velocity is known only at a finite number of points and, most importantly only an approximation of the real displacement of the landslide is known, due to error statistics in DInSAR (Sectio 1.2), it's necessary to approximate (3.1) numerically [7] [22] [23].

3.1.1 Finite Difference Approximation

The most simple way to approximate the derivative of a function in a point by means of the values of the function is the finite difference (FD) approximation.

Definition 6. *Given an univariate smooth function f and a point of the domain x, a* ***forward difference*** *approximation of stepsize h is simply the*

difference quotient

$$D_+(h) = \frac{f(x+h) - f(x)}{h},$$

similarly the ***backward difference*** *approximation is*

$$D_-(h) = \frac{f(x) - f(x-h)}{h}.$$

Finally, the ***central difference*** *approximation is*

$$D_0(h) = \frac{f(x+h) - f(x-h)}{2h}.$$

Since the derivative of a function at a point is defined as the limit of the difference quotient as the stepsize approaches to 0, the forward difference should be a fisrt approximation of the derivative. In particular the following result holds.

Theorem 3. *Given an univariate smooth function f and a point of the domain x, defining the* ***local truncation error*** *τ as the difference between the FD value and the true derivative value*

(i) if a forward difference is adopted then

$$\tau = \frac{f''(\xi)}{2}h, \quad \xi \in (x, x+h)$$

(ii) if a backward difference is adopted then

$$\tau = -\frac{f''(\xi)}{2}h, \quad \xi \in (x-h, x)$$

(iii) if a central difference is adopted then

$$\tau = \frac{f'''(\xi)}{6}h^2, \quad \xi \in (x-h, x+h)$$

Proof. In the case of (i)

$$\tau = D_+(h) - f'(x) = \frac{f(x+h) - f(x)}{h} - f'(x)$$

writing the Taylor expansion of f at $x+h$ with respect to x

$$f(x+h) = f(x) + f'(x)h + \sum_{n=2}^{\infty} \frac{f^{(n)}(x)}{n!} h^n$$

equivalently

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \sum_{n=2}^{\infty} \frac{f^{(n)}(x)}{n!} h^{n-1}$$

the local truncation error becomes

$$\tau = \sum_{n=2}^{\infty} \frac{f^{(n)}(x)}{n!} h^{n-1}$$

and mean value theorem provides the desired ξ.
Case (ii) is proved in the same way by applying the Taylor expansion of f at $x-h$ with respect to x. To prove (iii), consider

$$\tau = D_0(h) - f'(x) = \frac{f(x+h) - f(x-h)}{2h} - f'(x)$$

since

$$f(x+h) = f(x) + f'(x)h + \sum_{n=2}^{\infty} \frac{f^{(n)}(x)}{n!} h^n$$

$$f(x-h) = f(x) - f'(x)h + \sum_{n=2}^{\infty} (-1)^n \frac{f^{(n)}(x)}{n!} h^n$$

the local truncation error is

$$\tau = \frac{2f'(x)h + \sum_{n=2p+1} 2\frac{f^{(n)}(x)}{n!} h^n}{2h} - f'(x) = \frac{f'''(\xi)}{6} h^2, \quad \xi \in (x-h, x+h).$$

□

The previous theorem guarantees the consistency of the FD method in approximating the first derivative, in particular the forward and backward difference are first order convergent methods and the central difference is a second order convergent method.

For the second order derivative we can iteratively approximate it with a finite difference scheme, for example using a central difference

$$f''(x) \simeq \frac{f'(x+h) - f'(x-h)}{h} \simeq \frac{\frac{f(x+h)-f(x)}{h} - \frac{f(x)-f(x-h)}{h}}{h},$$

we can obtain the central difference approximation of the second order derivative as

$$f''(x) \simeq \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}. \tag{3.2}$$

With the same procedure in the proof of (iii), it can be shown that the local truncation is

$$\tau = \frac{f^{(4)}(\xi)}{12} h^2.$$

It's not difficult to generalize this method for multivariate functions, in particular the consistency result holds for each partial derivative of this function.

3.1.2 Thickness Linear System

In the case of Fels landslide, the superficial velocity field is known on a rectangular grid of 1 m resolution both along x and y, therefore it's natural to approximate the partial derivatives of the scaled thickness with stepsizes

$$h_x = h_y = 1 \text{ m}.$$

If we set N_x and N_y as the number of points along each x-direction and y-direction, the **discrete grid** is the set of all points$\{(x_i, y_j), i = 1, .., N_x \quad j = 1, .., N_y\}$. In our data

$$N_x = 2750, \quad N_y = 2050.$$

The vertical velocity scalar field is shown in Figure 3.1, negative values (dark regions) imply a downward displacement has occurred, and positive values (bright regions), an upward displacement.

The resolution of the grid is very high with respect to the size of the landslide and the magnitude of the horizontal velocity is generally very small, as a consequence to get an impression of the superficial horizontal velocity field we should plot a scaled version of it on a subgrid. In Figure 3.2 we plotted the horizontal velocity scaled by a factor of 10 on the subgrid of resolution 25 m.

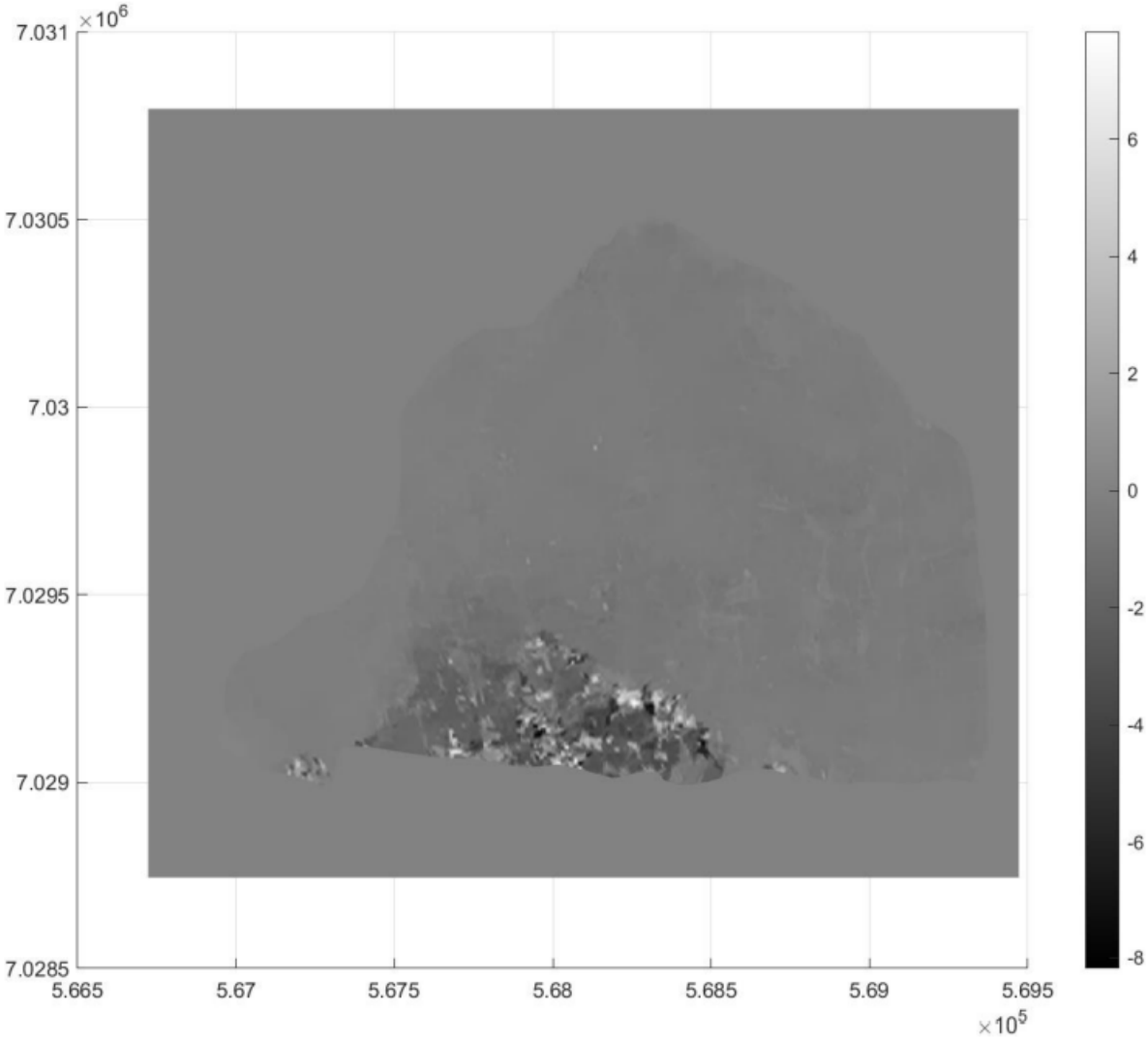


Figure 3.1: The vertical velocity field of Fels landslide on a 1 m resolution rectangular space. The unit is meter per year and data are derived from SAR ST (Section 1.2).

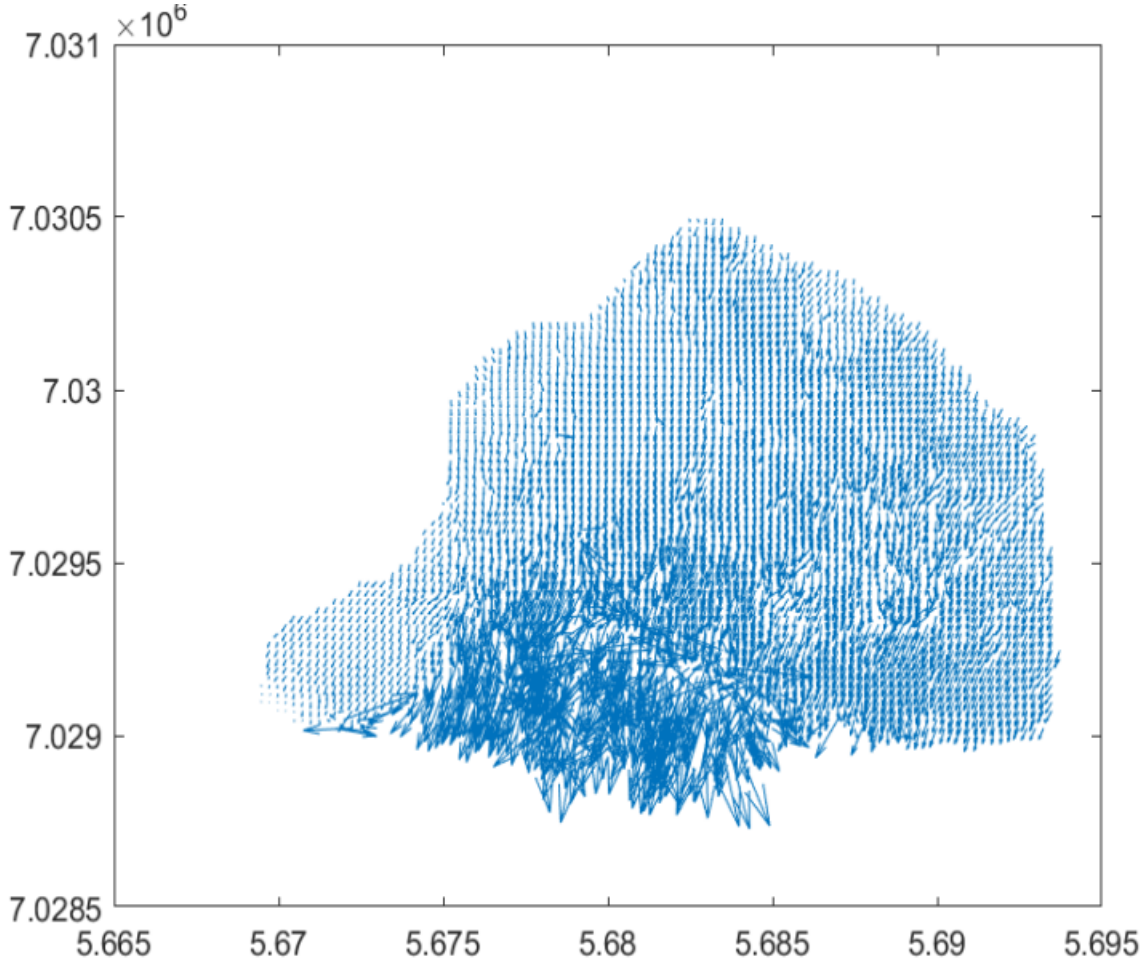


Figure 3.2: The horizontal velocity field of Fels landslide on a 25 m resolution subgrid. The measurement unit is meter per year and data are scaled by a factor of 10.

To simplify notation as much as we can, we introduce the following convention. For every function f, if we set

$$f(x_i, y_j) = f_{ij}$$

the equation (3.1) implies the following relation must hold at each grid point

(x_i, y_j)

$$(-v_z^{(surf)})_{ij} = (\frac{\partial h_\gamma}{\partial x})_{ij}(v_x^{(surf)})_{ij} + (\frac{\partial v_x^{(surf)}}{\partial x})_{ij}(h_\gamma)_{ij} + (\frac{\partial h_\gamma}{\partial y})_{ij}(v_y^{(surf)})_{ij} + (\frac{\partial v_y^{(surf)}}{\partial y})_{ij}(h_\gamma)_{ij}$$

With the previous notation, each function f on the grid identifies a matrix $F = [f_{ij}]_{\{i=1..N_x, j=1..N_y\}}$ which is the transpose of the visually evident matrix of values of f on the grid (Figure 3.3).

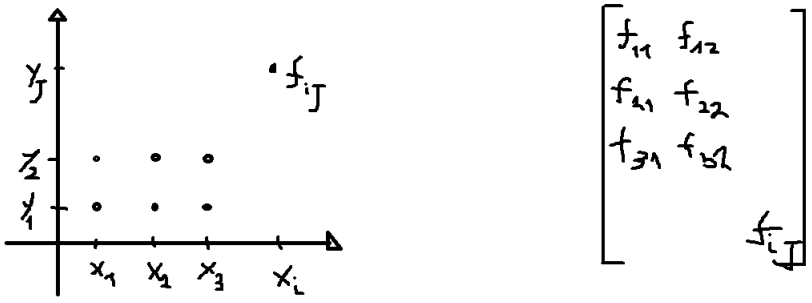


Figure 3.3: The discretization of a function and the associated matrix.

Now if we use a central difference approximation for both partial derivatives of the scaled thickness, the system of equations becomes

$$(-v_z^{(surf)})_{ij} = \frac{(h_\gamma)_{i+1j} - (h_\gamma)_{i-1j}}{2h_x}(v_x^{(surf)})_{ij} + \frac{(v_x^{(surf)})_{i+1j} - v_x^{(surf)})_{i-1j}}{2h_x}(h_\gamma)_{ij} + \frac{(h_\gamma)_{ij+1} - h_\gamma)_{ij-1}}{2h_y}(v_y^{(surf)})_{ij} + \frac{(v_y^{(surf)})_{ij+1} - (v_y^{(surf)})_{ij-1}}{2h_y}(h_\gamma)_{ij}.$$

Since $h_x = h_y = 1$

$$(-2v_z^{(surf)})_{ij} = ((h_\gamma)_{i+1j} - (h_\gamma)_{i-1j})(v_x^{(surf)})_{ij} + ((v_x^{(surf)})_{i+1j} - v_x^{(surf)})_{i-1j})(h_\gamma)_{ij} + ((h_\gamma)_{ij+1} - h_\gamma)_{ij-1})(v_y^{(surf)})_{ij} + ((v_y^{(surf)})_{ij+1} - (v_y^{(surf)})_{ij-1})(h_\gamma)_{ij}.$$

Writing explicitly with respect to the unknown scaled thickness h_γ and sorting first by y then by x, equivalently vectorizing the matrix identified by h_γ by column, we obtain the following equation

$$\begin{aligned}(-2v_z^{(surf)})_{ij} = &-(v_y^{(surf)})_{ij}(h_\gamma)_{ij-1} \\ &-(v_x^{(surf)})_{ij}(h_\gamma)_{i-1j} \\ &+(-(v_y^{(surf)})_{ij-1} - (v_x^{(surf)})_{i-1j} + (v_x^{(surf)})_{i+1j} + (v_y^{(surf)})_{ij+1})(h_\gamma)_{ij} \\ &+v_x^{(surf)})_{ij}(h_\gamma)_{i+1j} \\ &+(v_y^{(surf)})_{ij}(h_\gamma)_{ij+1}.\end{aligned}$$

Note that vectorizing the associated matrix by column means collecting rows of the visually evident matrix from bottom to top (figure 3.3).

In vectorial notation, for each grid point from bottom-left of the visually evident matrix to top-right

$$\forall j = 1..N_y \quad \forall i = 1..N_x$$

if we set $k = (j-1) + i$, by introducing b as the vector $b = [b_k]_{\{k=1..N_xN_y\}}$ with

$$b_k = (-2v_z^{(surf)})_{ij}$$

the system of linear equations becomes

$$Ah_\gamma = b, \tag{3.3}$$

where the matrix of coefficients A has a block-tridiagonal structure.

In particular it's tridiagonal with also (-N_x)-diagonal and N_x-diagonal different from zero vectors. The reason is, in each equation only 5 unknowns are involved which are the value of thickness at (x_i, y_j), the values at the previous and next points along x and y respectively. This particular choice is known as the **five-point stencil** structure. In particular, if we construct A row by row, on the diagonal, which will multiply by $(h_\gamma)_{ij}$, the coefficient is given by

$$-(v_y^{(surf)})_{ij-1} - (v_x^{(surf)})_{i-1j} + (v_x^{(surf)})_{i+1j} + (v_y^{(surf)})_{ij+1}$$

the previous and next coefficients are $-(v_x^{(surf)})_{ij}$ and $(v_x^{(surf)})_{ij}$. Finally at distance N_x from the diagonal there is on the left $-(v_y^{(surf)})_{ij}$ and $(v_y^{(surf)})_{ij}$ on the right.

3.2 The Regularized Problem

The system of linear equations to solve is huge: the unknown array has dimension 5 637 500. To make the problem more complicated, the matrix of coefficients, given completely by the horizontal velocity field, turns out to be numerically singular. In fact the matrix is very ill-conditioned. In order to give a physically meaningful result we have to apply results from Inverse Theory [8-13].

3.2.1 Fidelity Functional

Inverse problems are ubiquitous in nature: they have a central role in Imaging Science and Geophysics and they are becoming more and more important in many other fields of science and engineering. Example of applications include Computed Tomography, Astronomical Image Deblurring, Black Hole Imaging and Nuclear Magnetic Resonance data inversion for Petroleum Engineering, just to name a few. A formal definition of a linear inverse problem, which is so general to incorporate a myriad of engineering problems, can be given as follows.

Definition 7. *Let K be a bounded and linear operator between 2 Banach spaces X, Y*

$$K : X \to Y$$

*if $y^\dagger \in Y$, the **exact effect**, is the transformation of the **exact cause** $x^\dagger \in X$*

$$y^\dagger = Kx^\dagger$$

but only a noisy version of $y^\dagger$, the ***observed effect*** *y^δ, is known such that:*

$$\phi(x^\dagger, y^\delta) \leq \delta^2$$

for some metric $\phi : X \times Y \to \mathbb{R}$ and some ***noise level*** *$\delta > 0$,*
a ***linear inverse problem*** *is about to find $\hat{x}$, a guess or estimate of $x^\dagger$ based on observation y^δ.*

The operator K can represent an image formation process, for example a camera taking a picture of an object, or it can be a causal relationship between physical quantities. In our example (equation (2.4)), K is the differential operator linking the underlying thickness and the superficial vertical velocity of the landslide.

Almost always an inverse problem arising from application is an **ill-posed** problem, that is according to Hadamard, the solution may not exist, or if a such solution does exist it may not be unique, or a tiny change in observed effect may lead to a completely different guess. In the last case we say the problem is **unstable**. The general approach to deal with an ill-posed inverse problem is by **regularization**, namely we find and solve a new well-posed inverse problem which is, in a sense, "near" to the original one.

The classical example is given by the least square problem as a regularized version of an overdetermined linear system: instead of looking for the exact preimage of the vector b, we find a solution whose image is close enough to it in some metric. For example using Euclidean distance in $\mathbb{R}^n$, we can solve

$$Ax = b, \quad A \in \mathbb{R}^{m \times n}$$

with $m > n$, by solving

$$\min_{x \in \mathbb{R}^n} ||Ax - b||.$$

The previous approach can be generalized to the so-called variational methods.

Definition 8. *Given a linear inverse problem,* ***variational methods*** *are those seeking $\hat{x}$ as a minimizer of an* ***Energy functional*** *with parameter λ*

$$\min_{x \in X} J_\lambda(x), \quad J_\lambda : X \to \mathbb{R} \tag{3.4}$$

where the functional is made of 2 parts: a ***Fidelity term*** *$\phi(\cdot, y^\delta)$, which constrains the approximate solution to produce an effect similar to the observed one, and a* ***Regularization term*** *ψ which represents additional properties we think the exact solution might have, i.e.*

$$J_\lambda(x) = \phi(x, y^\delta) + \lambda\psi(x).$$

The fidelity term typically incorporates prior information about the data acquisition model, in particular about the noise affecting the observation. To clarify this claim let's recall the definition of a discrete linear inverse problem in a stochastic setting.

Definition 9. *Given a linear map between $\mathbb{R}^n$ and $\mathbb{R}^m$*

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

let $b^\dagger$ be the exact image of a point $x^\dagger$

$$b^\dagger = Ax^\dagger$$

if $\mathcal{N}$ is a random operator on $\mathbb{R}^m$, representing the noise affecting the measurements

$$\mathcal{N} : \Omega \times \mathbb{R}^m \to \mathbb{R}^m$$

assuming the observed effect b^δ is a realization of the random vector B

$$B = \mathcal{N}(b^\dagger)$$

such that

$$\phi(x^\dagger, b^\delta) \leq \delta^2$$

for some metric $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *and some noise level* $\delta > 0$,
a ***discrete linear inverse problem*** *is about to find an estimate* $\hat{x}$ *of* $x^\dagger$.

Definition 7 is concise and elegant, but it's not very of practical use. On the contrary, Definition 9 allows us to turn theoretical models of interest to something that can be represented and manipulated on a computer.

In the case of Fels landslide we will assume A is exactly the matrix in equation (3.3) and the vertical velocity is the realization of the noisy vertical velocity

$$b^\delta = b.$$

To be precise the noise is also affecting the operator K, but we will not deal with this aspect.

Theorem 4. *Given a discrete linear inverse problem, if the noise is* ***additive white Gaussian****, i.e.*

$$B = b^\dagger + N = Ax^\dagger + N,$$

where N *is a random vector having independent and normally distributed components* N_i *such that*

$$\forall i = 1, .., m \quad N_i \sim N(0, \sigma^2) \quad , \sigma \in \mathbb{R}$$

then the most likely guess is given by a variational method with the following fidelity term

$$\phi(\cdot; b^\delta) : \mathbb{R}^n \to \mathbb{R}$$
$$\phi(x; b^\delta) = \frac{1}{2}||Ax - b^\delta||^2$$

Proof. If we set U as the random vector representing the guess undergoing the process K, the most likely guess should maximize the probability of observing b^δ as the effect, in other words it should maximize the likelihood

$$\hat{x} \in arg \max_{u \in \mathbb{R}^m} \quad p_{(B|U)}(b^\delta, u).$$

Since the noise is additive white Gaussian the joint probability is the product of the marginals

$$p_{(B|U)}(b^\delta, u) = \prod_{i=1}^{m} p_{(B_i|U)}((b^\delta)_i, u)$$

where each component $B_i = (Ax)_i + N_i$ is a random variable, because N_i is a random variable. In particular since the latter is centered at the origin

$$B_i \sim N((Ax)_i, \sigma^2).$$

Now recalling the expression of the Gaussian distribution, the probability density of B given U can be written as

$$p_{(B|U)}(b^\delta, u) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{((b^\delta)_i - (Ax)_i)^2}{2\sigma^2}}$$

We can drop the positive constant in the maximization problem so

$$\hat{x} \in arg\max_{u \in \mathbb{R}^n} \quad e^{-\sum_{i=1}^{m} \frac{((b^\delta)_i - (Ax)_i)^2}{2\sigma^2}}$$

On the other hand, maximizing the previous function is equivalent to minimize the negative logarithm of the same function, so

$$\hat{x} \in arg\min_{u \in \mathbb{R}^n} \quad \sum_{i=1}^{m} \frac{((b^\delta)_i - (Ax)_i)^2}{2\sigma^2}.$$

Drop the positive constant σ^2 to conclude the proof. □

Assuming the noise affecting the measurements of the vertical velocity b in equation (3.3) is additive white Gaussian, in our case $m = n$, and by setting d as

$$d = N_x N_y = 5\ 637\ 500,$$

the fidelity term yielding the best guess is

$$\phi : \mathbb{R}^d \to \mathbb{R}$$
$$\phi(h_\gamma; b) = \frac{1}{2}||Ah_\gamma - b||^2$$

3.2.2 Regularization Functional

Given a general linear inverse problem, the additional information on the solution, represented by the regularization term, is usually a particular functional space it belongs to. This information can be for example its differentiability or, in the discrete case, its sparsity.

The classical regularization term is the **Tikhonov regularizer**, dating back in the 1960s [21]

$$\psi : \mathbb{R}^n \to \mathbb{R}$$
$$\psi(x) = \frac{1}{2}||Lx||^2,$$

where L is the discretization of some linear operator.

Booth et al. (2013) chose L to be the Laplacian operator in [6], which maps a generic bivariate function u onto

$$Lu = \Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Again, since the resolution in data is 1 m, if we use a central difference approximation, according the approximation (3.2), each second order partial derivative has discretization:

$$(\frac{\partial^2 u}{\partial x^2})_{ij} = u_{i+1j} - 2u_{ij} + u_{i-1j}$$
$$(\frac{\partial^2 u}{\partial x^2})_{ij} = u_{ij+1} - 2u_{ij} + u_{ij-1}$$

which imply

$$(\frac{\partial^2 u}{\partial x^2})_{ij} + (\frac{\partial^2 u}{\partial x^2})_{ij} = u_{ij-1} + u_{i-1j} - 4u_{ij} + u_{i+1j} + u_{ij+1}.$$

Therefore, in the same way we found the block-tridiagonal structure of the matrix of coefficients A in Subsection 3.1.2, the discretization matrix L representing the Laplacian is

$$\begin{bmatrix} T & I & & & \\ I & T & I & & \\ & I & T & I & \\ & & \ddots & \ddots & \ddots \\ & & & I & T \end{bmatrix}, \quad T = \begin{bmatrix} -4 & 1 & & & \\ 1 & -4 & 1 & & \\ & 1 & -4 & 1 & \\ & & \ddots & \ddots & \ddots \\ & & & 1 & -4 \end{bmatrix}$$

Figure 3.4: Picture from [7].

To recapitulate, an estimate of the underlying thickness of Fels landslide can be derived by solving the following minimization problem

$$\min_{h_\gamma \in \mathbb{R}^d} \frac{1}{2}||Ah_\gamma - b||^2 + \lambda \frac{1}{2}||Lh_\gamma||^2. \tag{3.5}$$

3.2.3 Implementation Details

A closer look at optimization problem (3.5) tells us the energy functional is quadratic, in particular it has form

$$\min_{h_\gamma \in \mathbb{R}^d} \frac{1}{2} h_\gamma^T Q h_\gamma - \tilde{b}^T x,$$

where Q is a positive-semidefinite symmetric matrix and $\tilde{b}$ a vector in $\mathbb{R}^d$. Indeed, by calling $x = h_\gamma$, we have

$$\begin{aligned} \frac{1}{2}||Ax - b||^2 + \lambda \frac{1}{2}||Lx||^2 &= \frac{1}{2}(Ax - b)^T(Ax - b) + \lambda \frac{1}{2}(Lx)^T Lx \\ &= \frac{1}{2}(x^T A^T - b^T)(Ax - b) + \lambda \frac{1}{2} x^T L^T Lx \\ &= \frac{1}{2}(x^T A^T Ax - x^T A^T b - b^T Ax + b^T b) + \lambda \frac{1}{2} x^T L^T Lx \\ &= \frac{1}{2}(x^T A^T Ax - 2x^T A^T b + b^T b) + \lambda \frac{1}{2} x^T L^T Lx. \end{aligned}$$

Since the variable is x, dropping the constant, the energy function to minimize is

$$\begin{aligned}\frac{1}{2}(x^TA^TAx - 2x^TA^Tb) + \lambda\frac{1}{2}x^TL^TLx &= \frac{1}{2}x^TA^TAx + \lambda\frac{1}{2}x^TL^TLx - x^TA^Tb \\ &= \frac{1}{2}x^T(A^TA + \lambda L^TL)x - (A^Tb)^Tx\end{aligned}$$

therefore, the final computational problem to solve is

$$\min_{x\in\mathbb{R}^d} \frac{1}{2}x^T(A^TA + \lambda L^TL)x - (A^Tb)^Tx \quad s.t. \quad x \geq 0. \tag{3.6}$$

The computational effort to solve this problem is enormous ($d = 5\ 637\ 500$), Booth et al. proposed a **bootstrapping** approach in which they downsampled all the 1 m resolution data to a 20 m grid, resulting in 400 independent estimates of h_f in each cell. We did all our experiments with cells of size 10 m by 10 m.

Formally instead of minimizing the energy functional (3.5) we are minimizing an approximation of it. Indeed, writing explicitly the components

$$\begin{aligned}\frac{1}{2}||Ax - b||^2 + \lambda\frac{1}{2}||Lx||^2 &= \frac{1}{2}\sum_{i=1}^{d}(Ax - b)_i^2 + \lambda\frac{1}{2}\sum_{i=1}^{d}(Lx)_i^2 \\ &= \sum_{i=1}^{d}\frac{1}{2}(Ax - b)_i^2 + \lambda\frac{1}{2}(Lx)_i^2\end{aligned}$$

the minimum is

$$\begin{aligned}\min_{x\in\mathbb{R}^d}\frac{1}{2}||Ax - b||^2 + \lambda\frac{1}{2}||Lx||^2 &= \min_{x\in\mathbb{R}^d}\sum_{i=1}^{d}\frac{1}{2}(Ax - b)_i^2 + \lambda\frac{1}{2}(Lx)_i^2 \\ &= \min_{x\in\mathbb{R}^d}\sum_{I=\text{SubgridIndex}}\sum_{i\in I}\frac{1}{2}(Ax - b)_i^2 + \lambda\frac{1}{2}(Lx)_i^2\end{aligned}$$

On the other hand, if we consider a cell of size 10 m $\times$ 10 m having index I, we can solve the optimization subproblem

$$\min_{x\in\mathbb{R}^{100}}\frac{1}{2}||A^{(I)}x - b^{(I)}||^2 + \lambda\frac{1}{2}||L^{(I)}x||^2$$

where $A^{(I)}$ is built using data on the cell and the outer perimeter, except for the 4 corners, due to particular discretization scheme chosen (five-point stencil).

Furthermore, they allowed the regularization parameter to vary on each cell, i.e. in our notation

$$\min_{x\in\mathbb{R}^{100}} \frac{1}{2}||A^{(I)}x - b^{(I)}||^2 + \lambda^{(I)}\frac{1}{2}||L^{(I)}x||^2.$$

Therefore the real computational problem solved on each cell is

$$\min_{x\in\mathbb{R}^{100}} \frac{1}{2}x^T((A^{(I)})^T A^{(I)} + \lambda^{(I)}(L^{(I)})^T L^{(I)})x - ((A^{(I)})^T b^{(I)})^T x \quad s.t. \quad x \geq 0 \tag{3.7}$$

3.2.4 Regularization Parameter

To be precise, on each cell we have identified a 1-parameter family of optimization subproblems, since for each different value of parameter λ the associated minimization problem gives a different guess.

The regularization parameter represents a trade-off between the fidelity term and the regularization term, or equivalently the weight given to the additional property of the solution. Booth et al. (2013) showed in [6] that a too small value of λ produces a highly oscillatory failure surface, on the other hand a value too big of the λ yields a thickness model that is overly smooth. Therefore choosing the regularization parameter is a crucial step.

For a generic linear inverse problem as in Definition 7, if the noise level is known or can be estimated, an a posteriori choice of regularization parameter proposed by Morozov is the following: let $x^\delta(\lambda)$ be a minimizer of the Energy functional in (3.4) with a fixed value of λ, an acceptable regularization parameter should have **realized residual** $\phi(x^\delta(\lambda), y^\delta)$ compatible with the noise level δ, i.e.

$$\phi(x^\delta(\lambda), y^\delta) = c\delta^2$$

for some small constant $c > 1$. The method is known as the **Discrepancy Principle** and in the case of a Gaussian noise, it's straightforward to pick the best estimate among several computed solutions.

Theorem 5. *Given a discrete linear inverse problem, if the noise is additive white Gaussian with standard deviation σ and if a variational method is chosen with the fidelity term as*

$$\phi(x, y^\delta) = ||Ax - y^\delta||^2,$$

then applying discrepancy principle is equivalent to require

$$||Ax^\delta(\lambda) - y^\delta||^2 = c(m\sigma^2)$$

for some small constant $c > 1$.

Proof. The observed effect y^δ is given by $\nu \in \mathbb{R}^m$, a realization of the random vector N

$$y^\delta = Ax^\dagger + \nu$$

so the squared norm of the realized noise is

$$||\nu||^2 = ||Ax^\dagger - y^\delta||^2.$$

On the other hand, since the noise random vector N has independent identically distributed components, its population variance $\hat{\sigma}^2$ is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^m \nu_i^2}{m} = \frac{||\nu||^2}{m}.$$

The exact solution makes the following equation true

$$||Ax^\dagger - y^\delta||^2 = m\hat{\sigma}^2,$$

therefore a good guess should have residual proportional to standard deviation of the noise. □

The discrepancy principle has been proved to be quite reliable for image denoising and deblurring tasks, however in practical applications the noise level is often unknown.

There exists another rule for choosing regularization parameter λ which can yield results comparable with those achievable with the previous criterion without any knowledge on σ. The basic idea is to balance the fidelity functional and the regularization functional, a general approach known as **Balancing Principle** [11].

Given a linear inverse problem, if a variational method is adopted then the **value function** is defined as

$$\begin{aligned} &F : (0, \infty) \to \mathbb{R} \\ &F(\lambda) = \inf_{x \in X} J_\lambda(x). \end{aligned}$$

For a given constant γ, assuming F is always positive, if we set Φ_γ as

$$\begin{aligned} &\Phi_\gamma : (0, \infty) \to \mathbb{R} \\ &\Phi_\gamma(\lambda) = \frac{F(\lambda)^{1+\gamma}}{\lambda}, \end{aligned}$$

the method proposed in [11] finds an optimal regularization parameter λ by minimizing Φ_γ.

Theorem 6. *Given a linear inverse problem with a variational method and a constant $\gamma > 0$, if $\bar{\lambda}$ is a local minimizer of Φ_γ then the value function F is differentiable at $\bar{\lambda}$. Also, if the associated energy functional $J_{\bar{\lambda}}$ admits a local minimizer $x^\delta(\bar{\lambda})$, then this guess makes the following equality true*

$$\phi(x^\delta(\bar{\lambda}), y^\delta) = \gamma \bar{\lambda} \psi(x^\delta(\bar{\lambda}))$$

The theoretical justifications of the rule is given by the following a posteriori error estimate.

Theorem 7. *Given a linear inverse problem with a variational method and a constant $\gamma > 0$. Assume both X and Y are Hilbert spaces, the fidelity term is*

$$\phi(x, y^\delta) = \frac{1}{2}||Kx - y^\delta||^2$$

and the regularization term is

$$\psi(x) = \frac{1}{2}||x||^2.$$

Assuming also that the exact solution is such that

$$x^\dagger = (K^*K)^\mu w,$$

where $||w|| \leq \rho$, with $\mu \in (0, 1]$ and $\rho > 0$, if $\bar{\lambda}$ is an optimal regularization parameter determined by the rule Φ_γ, by setting $\bar{\delta}$ as

$$\bar{\delta} = ||Kx^\delta(\bar{\lambda}) - y^\delta||$$

then the following estimate holds

$$||x^\delta(\bar{\lambda}) - x^\dagger|| \leq C(\rho^{\frac{1}{1+2\mu}} + \frac{F(\delta^{\frac{2}{1+2\mu}})^{\frac{1+\gamma}{2}}}{F(\bar{\lambda})^{\frac{1+\gamma}{2}}})max(\delta, \bar{\delta})^{\frac{2\mu}{1+2\mu}}, \quad C > 0$$

An efficient numerical algorithm is also provided in [11].

Algorithm 1 $BalancingPrinciple(y^\delta, \phi(\cdot, y^\delta), \psi, \lambda_0, maxit, tol) \rightarrow (\lambda, x^\delta)$

1: $\lambda = \lambda_0$;
2: **for** $k = 1, .., maxit$ **do**
3: $\quad x^\delta \in arg\min_{x \in X} \phi(x, y^\delta) + \lambda\psi(x)$;
4: $\quad \lambda_{old} = \lambda$;
5: $\quad \lambda = \frac{1}{\gamma}\frac{\phi(x^\delta, y^\delta)}{\psi(x^\delta)}$;
6: $\quad$ **if** $(|\lambda - \lambda_{old}| <= tol)$ **then**
7: $\quad\quad break$;
8: $\quad$ **end if**
9: **end for**

We note that λ_0 is an initial regularization parameter. This algorithm has a practically very desirable monotone convergence, under reasonable assumptions.

Theorem 8. *Given a linear inverse problem with a variational method and a constant $\gamma > 0$. If there exists an interval $[\zeta_0, \zeta_1]$ such that*

(i) $\phi(x^\delta(\zeta_1)) > 0$

(ii) $\exists \lambda_b \in [\zeta_0, \zeta_1]$ *which is an optimal parameter with*
$D^{\pm}\Phi_\gamma(\lambda) < 0 \quad \forall \lambda \in [\zeta_0, \lambda_b)$ *and*
$D^{\pm}\Phi_\gamma(\lambda) > 0 \quad \forall \lambda \in (\lambda_b, \zeta_1]$,

if also the initial parameter $\lambda_0 \in [\zeta_0, \zeta_1]$ then the sequence $\{\Phi_\gamma(\lambda_k)\}_{k\in\mathbb{N}}$ generated by Algorithm 1 is monotonically decreasing and the sequence $\{\lambda_k\}_{k\in\mathbb{N}}$ converges to the local minimizer λ_b.

For our purposes, to solve each minimization subproblem (3.7), the regularization parameter on line 5 will be updated as follows

$$\lambda^{(I)} = c\frac{||A^{(I)}h_\gamma^{(I)} - b^{(I)}||^2}{||L^{(I)}h_\gamma^{(I)}||^2};$$

for an appropriate constant $c > 0$.

3.3 Unconstrained Optimization

To recapitulate, we have found the relation between input data and the desired output, i.e. the thickness equation. We have discretized it because we have only a finite number of measurements. The linear system obtained in this way cannot be solved directly since measurements inevitably contain noises, therefore a regularization technique is required. In the end, what we really have to do is to solve a sequence of optimization problems, to be able to give an estimate of the underlying thickness of a landslide and infer its failure geometry. Thus it's necessary to recall some classical optimization methods [14] [15] [8] which provide the basis more modern algorithms rely on.

3.3.1 First Order Condition

Definition 10. *Given a multivariate function*

$$f : \mathbb{R}^n \to \mathbb{R}$$

an ***optimization problem*** *is about to find a global minimizer or a local minimizer of f.*

In practical applications f can be the loss of accuracy for an artificial neural network, the energy function of a physical system or the function representing the risk undertaken by a financial investor. Of course, from another perspective, an optimization problem can viewed as finding the maximizers of a function: for example the performance of a network, the probability of physical configurations or the rate of return achievable by an investor.

Almost always a numerical method will provide us a local minimizer of a loss function. But if the function is convex, the local minimizer is also global as shown in the following result.

Theorem 9. *Given an optimization problem, if f is* ***convex*** *i.e.*

$$\forall x, y \in \mathbb{R}^n \quad f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

then a point is a local minimizer if and only if it's a global minimizer. As a consequence, for a convex function we will refer to the solution simply as a ***minimizer****.*
Moreover, if f is ***strictly convex****, namely the previous inequality holds strictly, there can be at most one minimizer.*

Proof. If a $\bar{x}$ is a global minimizer of a f, it means

$$\forall x \in \mathbb{R}^n \quad f(\bar{x}) \leq f(x)$$

since the definition of a local minimizer is that the previous inequality must hold true for some neighborhood of the point, $\bar{x}$ is a local minimizer if we

take the neighborhood as the entire domain $\mathbb{R}^n$.
On the other hand, if now $\bar{x}$ is a local minimizer, let's suppose that it is not a global minimizer. There must exist a point y at which the function has a lower value

$$\exists y \in \mathbb{R}^n : f(y) < f(\bar{x})$$

since f is convex the value of f along the segment $[\bar{x}, y]$ is less or equal than a linear combination of the values at the ending points

$$f((1-t)\bar{x} + ty) \leq (1-t)f(\bar{x}) + tf(y)$$

but from the previous inequality

$$f((1-t)\bar{x} + ty) < (1-t)f(\bar{x}) + tf(\bar{x}) = f(\bar{x}),$$

which means that there cannot exist any neighborhood such that $\bar{x}$ is a local minimizer, contradicting the hypothesis. Therefore we conclude $\bar{x}$ must be also a global minimizer.
The second part of the theorem is also proved by contradiction: if there were two minimizers $\bar{x}, \bar{y}$, since f is stricly convex, we would have

$$f(\frac{1}{2}\bar{x} + \frac{1}{2}\bar{y}) < \frac{1}{2}f(\bar{x}) + \frac{1}{2}f(\bar{y}) = f(\bar{x})$$

where $f(\bar{y}) = f(\bar{x})$ due to definition of global minimizer. □

For a generic function, a good starting point to find its local minimizers is by checking the stationarity.

Theorem 10 (Fermat's)**.** *Given an optimization problem, if f is continuously differentiable then all its local minimizers are stationary.*

Proof. Let $\bar{x}$ be a local minimizer, let's suppose that is it not **stationary** i.e. $\nabla f(\bar{x}) \neq 0$.
Then we can consider the following function

$$g : \mathbb{R}^n \to \mathbb{R}$$
$$g(x) = -\nabla f(\bar{x})^T \nabla f(x)$$

it follows that

$$g(\bar{x}) = -||\nabla f(\bar{x})||^2 < 0$$

since g is continuous, because f is C^1, there is a neighborhood of $\bar{x}$ such that the previous inequality holds for all points in this set. In particular it's true along the direction of the negative gradient

$$\exists T > 0 : \forall t \in [0, T] \quad g(\bar{x} + t(-\nabla f(\bar{x}))) < 0$$

on the other hand, using Taylor expansion of f at each point along this line we have $\forall t \in [0, T]$

$$f(\bar{x} + t(-\nabla f(\bar{x}))) = f(\bar{x}) + < \nabla f(\bar{x} + \tau(-t\nabla f(\bar{x}))), t(-\nabla f(\bar{x})) >$$

with some $\tau \in (0, 1)$. If we take a closer look at the right hand side

$$f(\bar{x}) + < \nabla f(\bar{x} + \tau(-t\nabla f(\bar{x}))), t(-\nabla f(\bar{x})) >= f(\bar{x}) + tg(\bar{x} + \tau t\nabla f(\bar{x}))$$

since g is negative along this segment, we have

$$f(\bar{x}) + tg(\bar{x} + \tau t\nabla f(\bar{x})) < f(\bar{x})$$

which implies

$$f(\bar{x} + t(-\nabla f(\bar{x}))) < f(\bar{x}),$$

contradicting the fact $\bar{x}$ is a local minimizer. □

In the case of a convex loss function, the previous necessary condition is also sufficient. To prove it, we have to give an equivalent characterization of convexity of a function.

Theorem 11. *Given f a multivariate function*

$$f : \mathbb{R}^n \to \mathbb{R}$$

which is continuously differentiable, it is convex if and only if

$$\forall x \in \mathbb{R}^n, \forall h \in \mathbb{R}^n \quad f(x + h) \geq f(x) + \frac{\partial f}{\partial h}(x)$$

equivalently, if and only if

$$\forall x, y \in \mathbb{R}^n, \quad f(y) \geq f(x) + (y - x)^T \nabla f(x)$$

Proof. Let x be a point and h a direction from x, the partial derivative of f along h at x is defined as

$$\frac{\partial f}{\partial h}(x) = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$$

if we set y as

$$y = x + h$$

the right hand side can be written as

$$\frac{f(x+th) - f(x)}{t} = \frac{f(x+t(y-x)) - f(x)}{t}$$

since f is convex

$$\frac{f(x+t(y-x)) - f(x)}{t} \leq \frac{(1-t)f(x) + tf(y)) - f(x)}{t} = f(y) - f(x).$$

Therefore

$$\frac{\partial f}{\partial h}(x) \leq f(y) - f(x) = f(x+h) - f(x).$$

□

Corollary 1. *Given an optimization problem, if f is continuously differentiable and convex then the set of its minimizers coincides with the set of its stationary points.*

Proof. Given a stationary point $\bar{x}$, let's suppose it is not a (global) minimizer. Then there is at least a point y

$$\exists y \in \mathbb{R}^n : f(y) < f(\bar{x})$$

but from the characterization theorem, if we set h as $h = y - \bar{x}$ we know that

$$f(y) \geq f(\bar{x}) + \frac{\partial f}{\partial h}(\bar{x})$$

since the directional derivative can be written in terms the gradient

$$\frac{\partial f}{\partial h}(\bar{x}) = h^T \nabla f(\bar{x})$$

the previous inequality implies

$$h^T \nabla f(\bar{x}) < 0$$

contradicting the fact that $\bar{x}$ is a stationary. □

The most important example of convex function to keep in mind is a positive-semidefinite quadratic function.

Definition 11. *Given a multivariate function f, it is* ***quadratic*** *if and only if there's a symmetric matrix* $Q \in \mathbb{R}^{n\times n}$ *such that*

$$\forall x \in \mathbb{R}^n \quad f(x) = \frac{1}{2}x^T Q x - b^T x + c$$

where $b \in \mathbb{R}^n, c \in \mathbb{R}$

It's known that when Q is not the zero matrix, the roots of f form a quadratic hypersurface in $\mathbb{R}^n$. In particular if $n = 2$, the set of all roots is a conic section, whist if $n = 3$, a quadric surface [24].

Theorem 12. *A quadratic function is convex if and only if Q is positive semidefinite. It is strictly convex if and only if Q is positive definite.*

The theorem is a direct consequence of the following lemma, since a straightforward computation shows that the Hessian of a quadratic function is the constant matrix Q.

Lemma 3. *Given a multivariate function f, if the Hessian is positive semidefinite then f is convex. If the Hessian is positive definite, then f is strictly convex.*

Proof. Thanks to Theorem 9 it sufficies to prove that

$$\forall x, y \in \mathbb{R}^n, \quad f(y) \geq f(x) + (y - x)^T \nabla f(x)$$

Let x,y be 2 points in $\mathbb{R}^n$, the Taylor expansion of f at y with respect to x is

$$f(y) = f(x) + (y - x)^T \nabla f(x) + \frac{1}{2}(y - x)^T H_f(x + \tau(y - x))(y - x)$$

where $\tau \in (0, 1)$, since the Hessian is positive semidefinite, we have

$$f(y) \geq f(x) + (y - x)^T \nabla f(x).$$

The strict convexity result is a porism of proof to Theorem 9 and the above, in particular it suffices to read every inequality strictly. □

As a consequence of Corollary 1, an optimization problem with a positive-definite quadratic loss function can be solved by solving a linear system.

Theorem 13. *Given an optimization problem, if f is quadratic with a positive definite Q then the unique (Theorem 9) minimizer is given by the solution to the following system of linear equations*

$$Qx = b$$

This is true because a straightforward computation shows the gradient of a quadratic function is

$$\nabla f(x) = Qx - b.$$

3.3.2 Second Order Condition

For a generic optimization problem, if the loss function is also twice-continuously differentiable then there's a further condition a local minimizer must fulfil.

Theorem 14. *Given an optimization problem, if f is C^2 then all its local minimizers have positive semidefinite Hessian matrix.*

Proof. If $\bar{x}$ is a local minimizer, let's suppose that the Hessian of f at this point is not positive semidefinite.
It means

$$\exists p \in \mathbb{R}^n \setminus \{0\} : p^T H_f(\bar{x}) p < 0.$$

If we set g as

$$g : \mathbb{R}^n \to \mathbb{R}$$
$$g(x) = p^T H_f(x) p,$$

by construction

$$g(\bar{x}) < 0.$$

Since g is continuous because f is twice-continuously differentiable

$$\exists T > 0 : \forall t \in [0, T] \quad g(\bar{x} + tp) < 0.$$

On the other hand using Taylor expansion of f along this segment

$$\forall t \in [0,T] \quad f(\bar{x}+tp) = f(\bar{x}) + < \nabla f(\bar{x}), tp > + < \frac{H_f(\bar{x}+\tau tp)tp}{2}, tp >$$

for some $\tau \in (0,1)$. The stationary of $\bar{x}$ implies the second addend is zero, so

$$f(\bar{x}+tp) = f(\bar{x}) + \frac{t^2}{2} g(\bar{x}+\tau tp)$$

and the previous inequality implies

$$f(\bar{x}+tp) < f(\bar{x}),$$

contradicting the fact that $\bar{x}$ is a local minimizer. □

If a point has positive definite Hessian, the first order condition and the second order condition are also sufficient conditions, in particular

Theorem 15. *Given an optimization problem with a twice-continuously differentiable loss function, if $\bar{x}$ is a stationary point with positive definite Hessian, then $\bar{x}$ is a strict local minimizer.*

Proof. We want to prove that there is a neighborhood of $\bar{x}$ on which

$$\exists \mathcal{N} \text{ neighborhood } \bar{x} : \forall x \in \mathcal{N} \quad f(\bar{x}) < f(x).$$

First of all, there exists a neighborhood on which each point is positive definite

$$\exists R > 0 : \forall x \in B_R(\bar{x}) \quad H_f(x) \succ 0.$$

If we set the neighborhood as the ball of radius R centered in $\bar{x}$

$$\mathcal{N} = B_R(\bar{x})$$

then for any point $y \in \mathcal{N}$

$$f(y) = f(\bar{x}) + < \nabla f(\bar{x}), y-x > + < \frac{H_f(\bar{x}+\tau(y-x))(y-x)}{2}, y-x >$$

the stationarity of $\bar{x}$ implies

$$\forall y \in \mathcal{N} \quad f(y) = f(\bar{x}) + \frac{1}{2}(y-x)^T H_f(\bar{x}+\tau(y-x))(y-x) > f(\bar{x})$$

since

$$||\bar{x}+\tau(y-x)-\bar{x}|| = \tau||y-x|| < ||y-x|| < R.$$

□

3.3.3 Scaled Gradient Descent

So far, we have seen properties a solution to an optimization problem must possess. From the algorithmic point of view, all classical methods generate a sequence of iterates $\{x_k\}_{k\geq 0}$ from a initial guess x_0 and they terminate when either no more progress can be made or when it seems that a solution point has been approximated with sufficient accuracy.

Since we are looking for a local minimizer, given x_k, the most intuitive idea to find the next iterate $x_{k+1} \in \mathbb{R}^n$ is choosing one such that the loss function f decreases

$$f(x_{k+1}) < f(x_k).$$

This class of optimization algorithms is known as **iterative descent methods**.

A large family of iterative descent algorithms adopt the **line search strategy**, namely a direction from the current iterate is chosen and the new iterate is sought along this direction

$$x_{k+1} = x_k + \alpha_k p_k,$$

where $p_k \in \mathbb{R}^n$ is the **search direction** and $\alpha_k > 0$ the **search stepsize**.

To guarantee the decrease in the loss function the line-search direction must be a descent direction.

Definition 12. *Given an optimization problem, let x_k be a point in the domain $\mathbb{R}^n$ and $p_k \in \mathbb{R}^n$ a direction at x_k, p_k is a **descent direction** if and only if the directional derivative of f at x_k is negative*

$$\frac{\partial f}{\partial p_k}(x_k) = p_k^T \nabla f(x_k) < 0$$

Example 1. *Given an optimization problem, let x_k be a point in the domain. If D_k is a positive definite symmetric matrix then by setting*

$$p_k = D_k(-\nabla f(x_k))$$

we can prove that p_k is a descent direction. This is true since

$$D_k(-\nabla f(x_k))^T \nabla f(x_k) = -\nabla f(x_k)^T D_k \nabla f(x_k) < 0.$$

The class of methods which use a such descent direction is known as **scaled gradient descent**

$$x_{k+1} = x_k + \alpha_k D_k(-\nabla f(x_k)).$$

A different choice of the matrix D_k gives arise to a different descent method: when $D_k = I$ is the identity matrix, we obtain the classical **gradient descent**. When $D_k = H_f(x_k)^{-1}$, provided the Hessian is positive-definite symmetric, the method is known as **Newton's method**.
Whist if $D_k = B_k^{-1}$, where the matrix B_k is an approximation of the Hessian, the method is known as **Quasi-Newton**: notable examples are Davidon-Fletcher-Powell method (DFP) and Brotden-Fletcher-Goldfarb-Shanno (BFGS) method.

The value of f along a generic line-search direction p_k is

$$f(x_k + \alpha_k p_k) = f(x_k) + \alpha_k p_k^T \nabla f(x_k) + \mathcal{O}(\alpha_k^2)$$

It follows that if p_k is a descent direction, then $f(x_k + \alpha_k p_k) < f(x_k)$ for all positive but sufficiently small values of α_k. As a consequence, if a descent direction is fixed a sufficiently small search stepsize will provide an iterate with smaller value of f.

Even if the decreasing condition is guaranteed, the reduction may be insufficient for the method to converge. A popular condition on the search stepsize to ensure a sufficient reduction is given by Armijo, namely α_k is required to be such that

$$f(x_k + \alpha_k p_k) < f(x_k) + c_1 \alpha_k p_k^T \nabla f(x_k), \quad c_1 \in (0, 1).$$

On the other hand, to rule out unacceptably short steps it's useful to consider a **curvature condition** provided by Wolf

$$p_k^T \nabla f(x_k + \alpha_k p_k) > c_2 p_k^T \nabla f(x_k), \quad c_2 \in (c_1, 1).$$

In the case of a positive-definite quadratic loss function, we know that there exists a unique (global and local) minimizer. Therefore a greedy approach will always provides us the solution. In particular, if we use an iterative descent algorithm with the steepest descent direction

$$p_k = -\nabla f(x_k) = -(Qx_k - b)$$

and the stepsize which provides us the most decrease (**exact line search**)

$$\alpha_k = arg \min_{\alpha \in \mathbb{R}} f(x_k - \alpha_k \nabla f(x_k))$$

then the descent algorithm giving us the minimizer of the quadratic function is the classical steepest descent method. The minimizer is exactly the solution of linear system in Theorem 13.

3.4 Constrained Optimization

The optimization problem (3.7) we are trying to solve has an additional constraint: physically speaking, thickness must be a non-negative quantity. This is an example of constrained optimization problem [14] [16].

Definition 13. *Given an optimization problem, it is* ***constrained*** *if and only if the solution is required to belong to a closed convex set* Ω *of* $\mathbb{R}^n$*.*

$$\min_{x \in \mathbb{R}^n} f(x) \quad s.t.\ (subject\ to) x \in \Omega$$

The set Ω is known as the **feasible set**. A **global minimizer** is a point $\bar{x} \in \Omega$ such that

$$\forall x \in \Omega \quad f(\bar{x}) \leq f(x)$$

and a **local minimizer** $\bar{x}$ becomes a point such that there exists a neighborhood, in the induced topology on Ω, such that the previous inequality holds

$$\exists B_r(\bar{x}) : \forall x \in B_r(\bar{x}) \cap \Omega \quad f(\bar{x}) \leq f(x).$$

Note that the Definition 13 is equivalent to

$$\min_{x \in \Omega} f(x).$$

As a consequence, the unconstrained problem (Definition 10) can be seen as a particular constrained case, with $\Omega = \mathbb{R}^n$. In this perspective, many of the previous results can be generalized. For example, to a convex function, there's no distiction between local minimizer and global minimizer.

Porism 1. *Given a constrained optimization problem, if f is convex then a point is a local minimizer if and only if it's a global minimizer. If f is strictly convex, there can be at most one minimizer.*

Proof. The proof is identical to that of Theorem 7, where 1) we replace $\mathbb{R}^n$ with Ω, 2) the neighborhood is meant in the induced topology, 3) the segments $[\bar{x}, y], [\bar{x}, \bar{y}]$ lie completely in Ω due to its convexity. □

3.4.1 First Order Condition

Regarding the first order necessary condition, the stationarity of a point becomes a weaker property.

Theorem 16. *Given a constrained optimization problem, if f is continuously differentiable on Ω then any local minimizer $\bar{x}$ must make the following inequality true*

$$\forall x \in \Omega \quad (x - \bar{x})^T \nabla f(\bar{x}) \geq 0 \tag{3.8}$$

Proof. Let's suppose there exist a point $x \in \Omega$ such that

$$(x - \bar{x})^T \nabla f(\bar{x}) < 0$$

then we can consider the values of f along the segment $[\bar{x}, x]$

$$\phi : [0,1] \to \mathbb{R}$$
$$\phi(r) = f(\bar{x} + r(x - \bar{x})).$$

Note that ϕ is well-defined since Ω is convex. Moreover, since f is continuously differentiable, Taylor and mean value theorem imply

$$\phi(r) = f(\bar{x} + r(x - \bar{x})) = f(\bar{x}) + r(x - \bar{x})^T \nabla f(\bar{x} + \tau r(x - \bar{x})),$$

where $\tau = \tau(r) \in (0, 1)$.
Since the gradient of f is continuous, the previous inequality guarantees that for a sufficiently small value of r

$$(x - \bar{x})^T \nabla f(\bar{x} + \tau r(x - \bar{x})) < 0.$$

As a consequence, we have

$$f(\bar{x} + r(x - \bar{x})) = f(\bar{x}) + r(x - \bar{x})^T \nabla f(\bar{x} + \tau r(x - \bar{x})) < f(\bar{x})$$

which contradicts the hypothesis $\bar{x}$ is a local minimizer. □

When solving a constrained optimization problem, a point satisfying the condition (3.6) is said to be **stationary**.

Example 2. *If the feasible set is the non-negative orthant*

$$\Omega = \{x \in \mathbb{R}^n | \forall i = 1, .., n \quad x_i \geq 0\}$$

a local minimizer must be such that

$$\forall x \geq 0 \quad \sum_{i=1}^{n} (x_i - \bar{x}_i) \frac{\partial f}{\partial x_i}(\bar{x}) \geq 0,$$

In particular, by choosing $\forall i = 1, .., n \quad x = \bar{x} + e_i$ *we have* $\frac{\partial f}{\partial x_i}(\bar{x}) \geq 0$, *similarly* $\forall i : \bar{x}_i > 0$ *by setting* $x = \bar{x} - e_i$ *we conclude that corresponding partial derivative must be zero*

$$\forall i : \bar{x}_i > 0 \quad \frac{\partial f}{\partial x_i}(\bar{x}) = 0.$$

Just like the unconstrained case, stationarity is also a sufficient condition for convex functions.

Corollary 2. *Given a constrained optimization problem, if f is continuously differentiable and convex, a point is a minimizer if and only if condition (3.6) holds.*

The proof is a direct consequence of the following porism of Theorem 9.

Porism 2. *Given a function on Ω, a closed convex set of $\mathbb{R}^n$*

$$f : \Omega \to \mathbb{R}$$

if f is continuously differentiable then f is convex if and only if

$$\forall x, y \in \Omega \quad f(y) \geq f(x) + (y - x)^T \nabla f(x)$$

Proof. (of Corollary 2)
If $\bar{x}$ is stationary

$$\forall x \in \Omega \quad (x - \bar{x})^T \nabla f(\bar{x}) \geq 0$$

the convexity of f and Porism 2 imply

$$\forall x \in \Omega \quad f(x) \geq f(\bar{x}) + (x - \bar{x})^T \nabla f(\bar{x}) \geq f(\bar{x})$$

which prove that $\bar{x}$ is a (global) minimizer. □

3.4.2 Scaled Gradient Projection

There is a great variety of algorithms to solve a constrained optimization problem, for our purposes, we will recall those which can be viewed as constrained versions of the unconstrained descent algorithms.

Definition 14. *Given a constrained optimization problem, let x_k be a point in Ω and p_k a direction at x_k. p_k is a **feasible direction** if only if for any sufficient small $\lambda_k > 0$*

$$x_k + \lambda_k p_k \in \Omega.$$

Since Ω is convex, a feasible direction at a point x_k is of the form

$$p_k = \gamma(y_k - x_k)$$

with $y_k \in \Omega$ and $\gamma > 0$.

A **feasible descent method** is an iterative descent method with line search strategy, where the descent direction is also feasible. Namely, the iteration has form

$$x_{k+1} = x_k + \lambda_k(y_k - x_k)$$

with

$$(y_k - x_k)^T \nabla f(x_k) < 0.$$

If a such y_k doesn't exist, the method returns x_k.

The search stepsize of a feasible descent algorithm must be sufficiently small to also guarantee the next iterate belongs to Ω. Since Ω is convex, any $\lambda_k \leq 1$ will be enough. A class of feasible direction methods, which set y_k as the projection of some point, is given by the **Scaled Gradient Projection** methods (SGP) [17].

We need to introduce the projection operator first.

Definition 15. *Given a constrained optimization problem, if $B_k \in \mathbb{R}^{n\times n}$ is a positive-definite symmetric matrix, it can be seen as the matrix associated with a metric tensor which induces a norm*

$$||\cdot||_{B_k} : \mathbb{R}^n \to \mathbb{R}$$
$$||\xi||_{B_k} = \sqrt{\xi^T B_k \xi}.$$

*Let $z_k \in \mathbb{R}^n$, we define the **projection** of z_k on Ω with respect to the metric B_k as*

$$P_{\Omega,B_k}(z_k) = arg\min_{y\in\Omega} ||y - z_k||^2_{B_k}$$

According to the definition

$$P_{\Omega,B_k}(z_k) = arg\min_{y\in\Omega}(y-z_k)^T B_k(y-z_k),$$

since

$$(y-z_k)^T B_k(y-z_k) = y^T B_k y - 2y^T B_k z_k + z_k^T B_k z_k$$

dropping the last addend, which is constant with respect to y, the projection operator has form

$$P_{\Omega,B_k}(z_k) = arg\min_{y\in\Omega}\frac{1}{2}y^T B_k y - y^T B_k z_k.$$

The idea behind this class of methods is based on the following necessary condition of a local minimizer: the projection onto the feasible set in a metric of any displacement along the descent direction given by inverse of the metric is the stationary point itself.

Theorem 17. *Given a constrained optimization problem, if f is continuously differentiable and $\bar{x}$ is a local minimizer then for any positive scalar α_k, for any positive definite symmetric matrix D_k, $\bar{x}$ must be a fixed point of the following operator*

$$T_{\alpha_k,D_k} : \Omega \to \mathbb{R}^n$$
$$T_{\alpha_k,D_k}(x) = P_{\Omega,D_k^{-1}}(x + \alpha_k D_k(-\nabla f(x))).$$

It's tempting to implement a fixed point algorithm

$$x_{k+1} = x_k + \lambda_k(y_k - x_k)$$

with

$$y_k = T_{\alpha_k,D_k}(x_k).$$

But what guarantees the decrease in the loss function of the optimization problem? The answer is provided by the following result, also proved by Bonettini et al. (2009) in [17].

Theorem 18. *Given a constrained optimization problem with f a continuously differentiable function. Let x_k be a point in Ω, α_k a positive scalar and D_k a positive definite symmetric matrix such that*

$$||D_k||, ||D_k^{-1}|| \leq L, \quad L > 1.$$

Supposing x_k is not a stationary point, then $T_{\alpha_k, D_k}(x_k) - x_k$ is a descent direction at x_k.

Bonettini et al. (2009) proposed the following algorithm.

Algorithm 2 $SGP(f, x_0, M \in \mathbb{N}, c_1, \theta \in (0,1), maxit, tol) \rightarrow x$

$x = x_0$;
$fx = f(x)$;
for $k = 1, .., maxit$ **do**
    $fx = [fx; f(x)]$;
    $\alpha = StepsizeSelection$;
    $D = ScalingMatrixSelection$;
    $y = P_{\Omega, D^{-1}}(x + \alpha D(-\nabla f(x)))$;
    $d = y - x$;
    if $(||d|| < tol)$ **then**
        $break$;
    else
        $\lambda = 1$;
        $f_{max} = max(fx(max(1, k - (M-1)) : k))$;
        while $(f(x + \lambda d) > f_{max} + c_1 \lambda d^T \nabla f(x))$ **do**
            $\lambda = \theta \lambda$;
        end while
        $x_{old} = x$; ▷ useful for StepsizeSelection
        $x = x + \lambda d$;
    end if
end for

In particular in choosing the search stepsize λ_k, a generalization of Armijo's rule is adopted. The non monotone line-search strategy implemented on lines 12-16 ensure the new value of the loss function is lower than the maximum on the last M-iterations.

The convergence result is given by the following theorem.

Theorem 19. *Given a constrained optimization problem with f a continuously differentiable function. If SGP is adopted with D_k as in Theorem 18 and an initial guess such that the set $\{x \in \Omega | f(x) \leq f(x_0)\}$, then every accumulation point of the sequence $\{x_k\}_{k\in\mathbb{N}}$ is a stationary point.*

Similarly to the unconstrained case, a different choice in the scaling matrix D_k on line 6 of Algorithm 2 yields a different method. When for each iteration the scaling matrix is just the identity

$$D_k = I \quad \forall k$$

we obtain the classical **gradient projection** method. In this case the update of x becomes

$$x_{k+1} = x_k + \lambda_k(y_k - x_k), \quad y_k = [x_k - \alpha_k \nabla f(x_k)]^+$$

where $[\cdot]^+$ represents the projection operator on Ω with respect to the Euclidean metric.

The simplest scaling matrix which is not the identity is a diagonal one

$$D_k = diag(d_1^{[k]}, .., d_n^{[k]}).$$

A diagonal scaling matrix allows one to make the projection on line 7 of Algorithm 2 a non-excessively expensive task. But the convergence rate of the diagonally scaled gradient projection is often unacceptably slow.

An efficient choice of a diagonal scaling matrix D_k is an approximation the inverse of the Hessian matrix at x_k. For example by requiring

$$d_i^{[k]} \approx \frac{1}{\frac{\partial^2 f}{\partial x_i^2}(x_k)}$$

an updating rule could be

$$d_i^{[k]} = min(L, max(\frac{1}{L}, \frac{1}{\frac{\partial^2 f}{\partial x_i^2}(x_k)}))$$

or

$$d_i^{[k]} = min(L, max(\frac{1}{L}, (x_k)_i)).$$

Regarding the choice of α_k on line 3, several stepsize updating strategies have been devised to accelerate the slow convergence exhibited in most cases by standard gradient methods. Numerical experiments on randomly generated, library and real-life test problems have shown the class updating rules, originally proposed by Barzilai and Borwein (BB) [25], has a remarkable convergence rate improvement.

If we set the matrix S_k as function of stepsize α_k

$$S_k(\alpha_k) = (\alpha_k D_k)^{-1}$$

as an approximation of the Hessian $H_f(x_k)$, we can derive two BB updating rules for α_k by forcing quasi-Newton properties on S_k

$$\alpha_k^{(BB1)} = arg \min_{\alpha \in \mathbb{R}} ||S_k(\alpha)\Delta x^{(k-1)} - \Delta g^{(k-1)}||$$

and

$$\alpha_k^{(BB2)} = arg \min_{\alpha \in \mathbb{R}} ||\Delta x^{(k-1)} - S_k(\alpha)^{-1}\Delta g^{(k-1)}||$$

with $\Delta x^{(k-1)} = x_k - x_{k-1}$ and $\Delta g^{(k-1)} = \nabla f(x_k) - \nabla f(x_{k-1})$.

In this way, the stepsizes

$$\alpha_k^{(1)} = \frac{(\Delta x^{(k-1)})^T D_k^{-1} D_k^{-1} \Delta x^{(k-1)}}{\Delta x^{(k-1)} D_k^{-1} \Delta g^{(k-1)}}$$

and

$$\alpha_k^{(2)} = \frac{(\Delta x^{(k-1)})^T D_k \Delta g^{(k-1)}}{\Delta g^{(k-1)} D_k D_k \Delta g^{(k-1)}}$$

are obtained, that reduce to the standard BB rules in the case of non-scaled gradient methods, namely if $D_k = I$

$$\alpha_k^{(1)} = \frac{(\Delta x^{(k-1)})^T \Delta x^{(k-1)}}{\Delta x^{(k-1)} \Delta g^{(k-1)}}$$

and

$$\alpha_k^{(2)} = \frac{(\Delta x^{(k-1)})^T \Delta g^{(k-1)}}{\Delta g^{(k-1)} \Delta g^{(k-1)}}.$$

At this point the authors of [17] proposed a stepsize updating rule which adaptively alternates the values provided by the previous equations. The rule decides the alternation between 2 different selection strategies by means of the variable threshold τ_k as shown in the following algorithm.

Algorithm 3 *StepsizeSelection*$(\alpha_{min}, \alpha_{max}) \rightarrow \alpha$

if (k==1) **then**
  $\alpha \in [\alpha_{min}, \alpha_{max}]$;
  $\tau \in (0, 1)$;
  $\alpha_2 = [\quad]$;
  $M_\alpha \in \mathbb{N}$;
else
  $dx = x - x_{old}$;
  $dg = \nabla f(x) - \nabla f(x_{old})$;
  if $(dx^T D^{-1} dg <= 0)$ **then**
    $\alpha_1 = \alpha_{max}$;
  else
    $\alpha_1 = max(\alpha_{min}, min(\alpha_{max}, \frac{dx^T D^{-1} D^{-1} dx}{dx^T D^{-1} dg}))$;
  end if
  if $(dx^T D_k dg <= 0)$ **then**
    $\alpha_2 = [\alpha_2; \alpha_{max}]$;
  else
    $\alpha_2 = [\alpha_2; max(\alpha_{min}, min(\alpha_{max}, \frac{dx^T D_k dg}{dg^T D_k D_k dg}))]$;
  end if
  $\tau_{old} = \tau$;
  if $(\frac{\alpha_2}{\alpha_1} > \tau)$ **then**
    $\alpha = \alpha_1$;
    $\tau = 1.1 * \tau$;
  else
    $\alpha = min(\alpha_2(max(1, k - M_\alpha) : k - 1))$;
    $\tau = 0.9 * \tau$;
  end if
end if

3.4.3 2-Metric Projection

In the following we will set the feasible set Ω as the non-negative orthant.

$$\Omega = \{x \in \mathbb{R}^n | x_i \geq 0 \quad \forall i = 1, .., n\}$$

The principal drawback of the SGP is that the projection on line 7, with a non diagonal scaling matrix, can be excessively complicated.

A natural and simple adaptation of unconstrained Newton-like methods is the **2-metric projection** method, presented in [16], which has the general iterate given by

$$x_{k+1} = [x_k + \alpha_k D_k(-\nabla f(x_k))]^+$$

where

$$[\cdot]^+ \mathbb{R}^n \to \mathbb{R}^n$$
$$([x]^+)_i = max(0, x_i), \quad \forall i = 1, .., n.$$

The fundamental difficulty here is that an arbitrary matrix D_k will not necessarily yield a descent direction.

It turns out, however, that there is a class of non-diagonal matrices for which descent is guaranteed. This class is sufficiently wide to allow superlinear convergence when D_k properly embodies second derivative information. To prove this claim, first we need a weaker notion of diagonal matrices.

Definition 16. *Given a square matrix $D \in \mathbb{R}^{n \times n}$ and a set of indices $I \subseteq \{1, .., n\}$, D is **diagonal with respect to** I if and only if*

$$\forall i \in I \quad \forall j \neq i \quad D_{ij} = 0$$

and

$$\forall j \in I \quad \forall i \neq j \quad D_{ij} = 0.$$

Similarly to Theorem 17, the idea of this new algorithm comes from the following equivalent condition of stationarity.

Theorem 20. *Given a constrained optimization problem on the non negative orthant, if f is continuously differentiable and $\bar{x}$ is a local minimizer then, defining $I^+(\bar{x})$ as*

$$I^+(\bar{x}) = \{i = 1, .., n | \bar{x}_i = 0 \wedge \frac{\partial f}{\partial x_i}(\bar{x}) > 0\}$$

then for any positive scalar α_k and for every positive definite symmetric matrix D_k, which is diagonal with respect to $I^+(\bar{x})$, $\bar{x}$ must be a fixed point of the following operator

$$T_{\alpha_k, D_k} : \Omega \to \mathbb{R}^n$$
$$T_{\alpha_k, D_k}(x) = [x + \alpha_k D_k(-\nabla f(x))]^+.$$

Proof. We will prove the condition is equivalent to stationarity. First, if $\bar{x}$ is a stationary point, let α_k be a positive scalar and D_k a positive definite symmetric matrix diagonal with respect to $I^+(\bar{x})$. We can assume, without loss of generality, that

$$I^+(\bar{x}) = \{r + 1, .., n\}$$

such that

$$D_k = \begin{pmatrix} \bar{D} & 0 & 0 & \cdots & 0 \\ 0 & d_{r+1} & 0 & \cdots & 0 \\ 0 & 0 & d_{r+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{pmatrix}$$

Figure 3.5: Picture from [16].

where $\bar{D}$ is positive definite symmetric and $d_{r+1}, .., d_n > 0$. From the stationarity of $\bar{x}$ (Example 2) and the definition of $I^+(\bar{x})$, we have

(i) $\forall i = 1, .., r \quad \frac{\partial f}{\partial x_i}(\bar{x}) = 0$

(ii) $\forall i = r+1, .., n \quad \frac{\partial f}{\partial x_i}(\bar{x}) > 0$

so the first r components of $T_{\alpha_k, D_k}(\bar{x})$ are exactly those of $\bar{x}$ since no displacement has been made along these directions. The last $n-r$ components are also the same since

$$\forall i = r+1, .., n \quad \bar{x}_i + \alpha_k d_i(-\frac{\partial f}{\partial x_i}(\bar{x})) < \bar{x}_i = 0$$

and the projection of a negative component is just $0 = \bar{x}_i$.
On the other hand, let the condition be true we show $\bar{x}$ is stationary. For a given matrix D_k and scalar $\alpha_k > 0$, by setting

$$q = D_k \nabla f(\bar{x})$$

we must have

(i) $q_i = 0 \quad \forall i = 1, .., r$ with $\bar{x}_i > 0$

(ii) $q_i \geq 0 \quad \forall i = r+1, .., n$ with $\bar{x}_i = 0$

since

$$T_{\alpha_k, D_k}(\bar{x}) = [\bar{x} - \alpha_k q]^+$$

Now by the definition of $I^+(\bar{x})$, we have that if $x_i = 0$ and $i \notin I^+(\bar{x})$, then $\frac{\partial f}{\partial x_i}(\bar{x}) \leq 0$. This together with the relations above, imply that

$$\sum_{i=1}^{r} \frac{\partial f}{\partial x_i}(\bar{x}) \leq 0.$$

Since

$$\begin{bmatrix} q_1 \\ . \\ . \\ q_r \end{bmatrix} = \bar{D} \begin{bmatrix} \frac{\partial f}{\partial x_1}(\bar{x}) \\ . \\ . \\ \frac{\partial f}{\partial x_r}(\bar{x}) \end{bmatrix}$$

while $\bar{D}$ is positive definite, we also have

$$\sum_{i=1}^{r} q_i \frac{\partial f}{\partial x_i}(\bar{x}) \geq 0$$

and it follows that

$$q_i = \frac{\partial f}{\partial x_i}(\bar{x}) = 0 \quad \forall i = 1, .., r.$$

Since

$$\forall i = r+1, .., n \quad \frac{\partial f}{\partial x_i}(\bar{x}) > 0 \wedge x_i = 0$$

we see that $\bar{x}$ is a stationary point. □

The decrease in the loss function is guaranteed by the following result.

Theorem 21. *Given a constrained optimization problem on the non negative orthant with f continuously differentiable function. Let x_k be a point in Ω which is not a stationary point. Then for every positive definite symmetric matrix D_k which is diagonal with respect to $I^+(x_k)$, there exists a positive scalar α_k such that*

$$f([x_k + \alpha_k D_k(-\nabla f(x_k))]^+) < f(x_k)$$

Proof. We'll assume D_k has the same structure as in the previous proof, i.e.

$$\forall i = r+1, .., n \quad (x_k)_i = 0 \wedge \frac{\partial f}{\partial x_i}(x_k) > 0$$

as a consequence

$$\forall i = r+1, .., n \forall \alpha_k > 0 \quad ([x_k - \alpha_k \nabla f(x_k)]^+)_i = (x_k)_i = 0.$$

Consider the sets of indices

$$I_1 = \{i = 1, .., r | (x_k)_i > 0 \vee ((x_k)_i = 0 \wedge (q_k)_i < 0)\}$$

with $q_k = D_k \nabla f(x_k)$ and

$$I_2 = \{i = 1, .., r | (x_k)_i = 0 \wedge (q_k)_i \geq 0\}.$$

Let $\bar{\alpha}$ be

$$\bar{\alpha} = sup\{\alpha | (x_k)_i - \alpha q_k \geq 0 \quad \forall i \in I_1\},$$

if we define the direction p_k as

$$(p_k)_i = \begin{cases} -(q_k)_i & \text{if } i \in I_1 \\ 0 & \text{if } i \in I_2 \vee i \in I^+(x_k) \end{cases} \tag{3.9}$$

then

$$\forall \alpha_k \in (0, \bar{\alpha}) \quad T_{\alpha_k, D_k}(x_k) = x_k + \alpha_k p_k.$$

In view of definition of I_2 and $I^+(x_k)$ we obtain

$$\frac{\partial f}{\partial x_i}(x_k) \leq 0 \quad \forall i \in I_2$$

and hence

$$\sum_{i \in I_2} \frac{\partial f}{\partial x_i}(x_k) q_i \leq 0.$$

Therefore, from the definition of p_k

$$p_k^T \nabla f(x_k) = \sum_{i \in I_1} -(q_k)_i \frac{\partial f}{\partial x_i}(x_k) \leq \sum_{i=1}^{r} -(q_k)_i \frac{\partial f}{\partial x_i}(x_k).$$

Since x_k is not stationary, from the previous theorem, we must have $T_{\alpha_k, D_k}(x_k) \neq x_k$ for some $\alpha_k > 0$ and hence also, $(q_k)_i \neq 0$ for some $i \in \{1, .., r\}$. Since $\bar{D}_k$ is is positive definite, we have

$$\sum_{i=1}^{r} -(q_k)_i \frac{\partial f}{\partial x_i}(x_k) < 0$$

proving that p_k is a descent direction. □

3.4.4 Newton Projection

To guarantee convergence, the implementation should be done more carefully. The reason is that the set $I^+(x_k)$ exhibits an undesirable discontinuity at the boundary of the constraint set, whereby given a sequence (x_k) of interior points that converges to a boundary point $\bar{x}$, the set $I^+(x_k)$ may be strictly smaller than the set $I^+(\bar{x})$. This causes difficulties in proving convergence of the algorithm and may have an adverse effect on its rate of convergence.

To bypass these difficulties one may add to the former set the indices of those

variables $(x_k)_i$ that satisfy $\frac{\partial f}{\partial x_i}(x_k) > 0$ and are "near" zero, i.e. the set of indices to be considered is

$$I_\epsilon^+(x_k) = \{i = 1, .., n | 0 \leq (x_k)_i \leq \epsilon \wedge \frac{\partial f}{\partial x_i}(x_k) > 0\}$$

With such a modification and with a variation of the Armijo rule on the projection arc, one can prove a satisfactory convergence result.

It is also possible to construct Newton-like algorithms, where the diagonal portion of the matrix D_k consists of the inverse of the Hessian submatrix corresponding to the indices not in $I^+(x_k)$, and to show a quadratic convergence result under the appropriate assumptions.

Landi and Loli Piccolomini (2008) proposed in [18] a method of such form referred to as **Complete-Hessian Newton Projection Conjugate Gradient** (CNPCG). In particular they chose a different tolerance for each iteration, i.e. $\forall k$ they considered $I_{\epsilon_k}^+(x_k)$ with

$$\epsilon_k = min(\epsilon, w_k), \quad w_k = ||x_k - [x_k - \nabla f(x_k)]^+||$$

Moreover, instead of incorporating second order information, like in [26], by imposing

$$p_k : H_k^{(I)} p_k = -\nabla f(x_k)$$

where $H_k^{(I)}$ is the **reduced Hessian**

$$(H_k^I)_{ij} = \begin{cases} \delta_{ij} & \text{if } i \in I_{\epsilon_k}^+(x_k) \vee j \in I_{\epsilon_k}^+(x_k) \\ (H_f(x_k))_{ij} & \text{otherwise,} \end{cases} \tag{3.10}$$

they solved first the linear solution using all information available

$$d_k : H_f(x_k) d_k = -g_k^{(I)}$$

where $g_k^{(I)}$ is the **reduced gradient**

$$(g_k^I)_i = \begin{cases} 0 & \text{if } i \in I_{\epsilon_k}^+(x_k) \\ (\nabla f(x_k))_i & \text{otherwise.} \end{cases} \tag{3.11}$$

Only then, they consider the search direction p_k

$$(p_k)_i = \begin{cases} -(\nabla f(x_k))_i & \text{if } i \in I^+_{\epsilon_k}(x_k) \\ (d_k)_i & \text{otherwise.} \end{cases} \tag{3.12}$$

The stepsize α_k is computed with the variation of the Armijo rule discussed in [26]. In particular, α_k is the first natural power of $\frac{1}{2}$ such that

$$f(x_k) - f([x_k + (\frac{1}{2})^m p_k]^+) \geq$$
$$\eta(-(\frac{1}{2})^m \sum_{i \notin I^+_{\epsilon_k}(x_k)} (\nabla f(x_k))_i (p_k)_i + \sum_{i \in I^+_{\epsilon_k}(x_k)} (\nabla f(x_k))_i ((x_k)_i - ([x_k + (\frac{1}{2})^m p_k]^+)_i))$$

with $\eta \in (0, \frac{1}{2})$. The proposed algorithm is the following.

Algorithm 4 $CNPCG(f, x_0, \eta \in (0, \frac{1}{2}), maxit, tol) \rightarrow x$

1: $x = x_0$;
2: **for** $k = 1, .., maxit$ **do**
3: $\quad I = \{i = 1, .., n | 0 \leq x(i) \leq \epsilon \wedge \frac{\partial f}{\partial x_i}(x) > 0\}$;
4: $\quad g = \nabla f(x); g(I) = 0$;
5: $\quad d = CG(H_f(x), g)$; ▷ Coniugate Gradient
6: $\quad p = d$;
7: $\quad p(I) = g(I)$;
8: $\quad \alpha = StepSelection(\eta)$;
9: $\quad x = max(0, x - \alpha p)$
10: $\quad ToleranceChecking(tol)$;
11: **end for**

On line 5 the approximation solution is obtained by the Conjugate Gradient method. About the stopping criterion, let $g_k^{(proj)}$ be the **projected gradient**

$$(g_k^{(proj)})_i = \begin{cases} (\nabla f(x_k))_i & \text{if } (x_k)_i > 0 \vee ((x_k)_i = 0 \wedge (\nabla f(x_k))_i < 0) \\ 0 & \text{otherwise,} \end{cases} \tag{3.13}$$

the Algorithm 4 stops before reaching the maximum number of iterations either when the norm of the projected gradient is sufficiently close to 0

$$||g_k^{(proj)}|| \leq tol||g_0^{(proj)}||$$

or when the relative distance between 2 consecutive iterates has become less than a tolerance

$$\frac{||x_{k+1} - x_k||}{||x_{k+1}||} \leq tol.$$

The convergence of the algorithm is provided by the following result.

Theorem 22. *Given a constrained optimization problem on the non-negative orthant with f twice-continuously differentiable. If the CNPCG algorithm is applied and if for each iteration k*

$$\exists \mu_1, \mu_2 > 0 : \quad \forall \quad \xi \in \mathbb{R}^n \mu_1 ||\xi||^2 \leq \xi^T H_f(x_k) \xi \leq \mu_2 ||\xi||^2$$

then every limit point of the sequence generated by the method is a stationary point.

Chapter 4

Numerical Results

In this chapter we present the results obtained with the proposed inversion method with 4 different implementations and Fels landslide data.

The input data (Section 1.2) is a table with five different columns; each row identifies a grid point belonging to the 1 m resolution rectangular discrete grid used to discretize the free surface of the slide. In particular, the first and second columns give the spatial coordinate of the point and the remaining three columns are the Cartesian components of the velocity vector at this point. The data comprises 5 637 500 grid points.

We first implemented the exact method proposed by Booth et al. (2013) in [6], in particular the bootstrapping approach is adopted and the regularization parameter of each optimization subproblem is found by applying the discrepancy principle. Then, we improved this method by applying balancing principle (Subsection 3.2.4) in the second experiment. In the third and the fourth experiments we kept balancing principle as the criterion to determine the regularization parameter and we experimented the SGP (Subsection 3.4.2) and CNPCG (Subsection 3.4.4) respectively.

4.1 Numerical Experiment 1

In article [6] the constrained optimization problem is solved using the program CVX, a Matlab-based modeling system for convex optimization [27] [28]. The regularization parameter is found by applying the discrepancy principle where the authors of the article use the standard deviation of b as an approximation of the standard deviation of the noise.

Following their notation by setting $\lambda = \alpha^2$ we chose 10 evenly spaced values of α between 10^{-2} and 10, then on each cell we solved the the optimization problem using the CVX program saving both the actual minimizer vector and its mean value. We also saved on each cell the best value of α provided by the discrepancy principle (Subsection 3.2.4) as shown in Figure 4.3.

Let's consider first the intensity of the velocity (Figure 4.1). It's straightforward to see that the intensities on the south are much larger than those on the north, this is one of factors contributing to peak values found here.

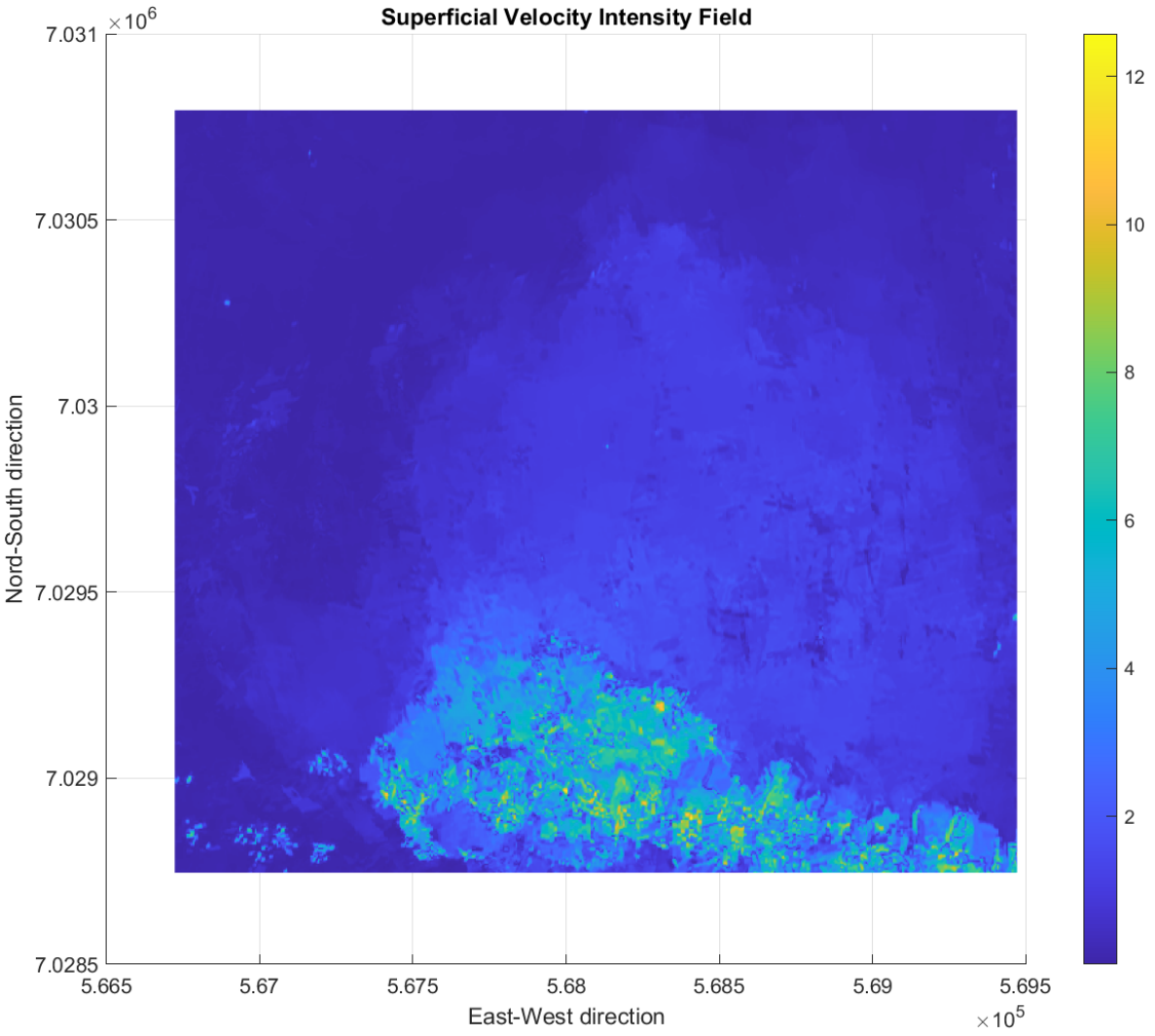


Figure 4.1: Superficial velocity intensity field. Unit of measurement: meter per year.

Since the most interesting data are away from the south and on the south the frontal moraine and glacier make measurements of landslide displacement particularly noisy, we decided to consider only values below a threshold to capture the general geometry. In particular for a value of $\gamma = 0.035$, the thickness inferred by the CVX method is the one shown in Figure 4.2. To be more precise in Figure 4.2 it's plotted the mean value which is constant on each cell, since the graph of the solution is exactly the same but with black

lines corresponding to smaller values on the northern edge of each cell.

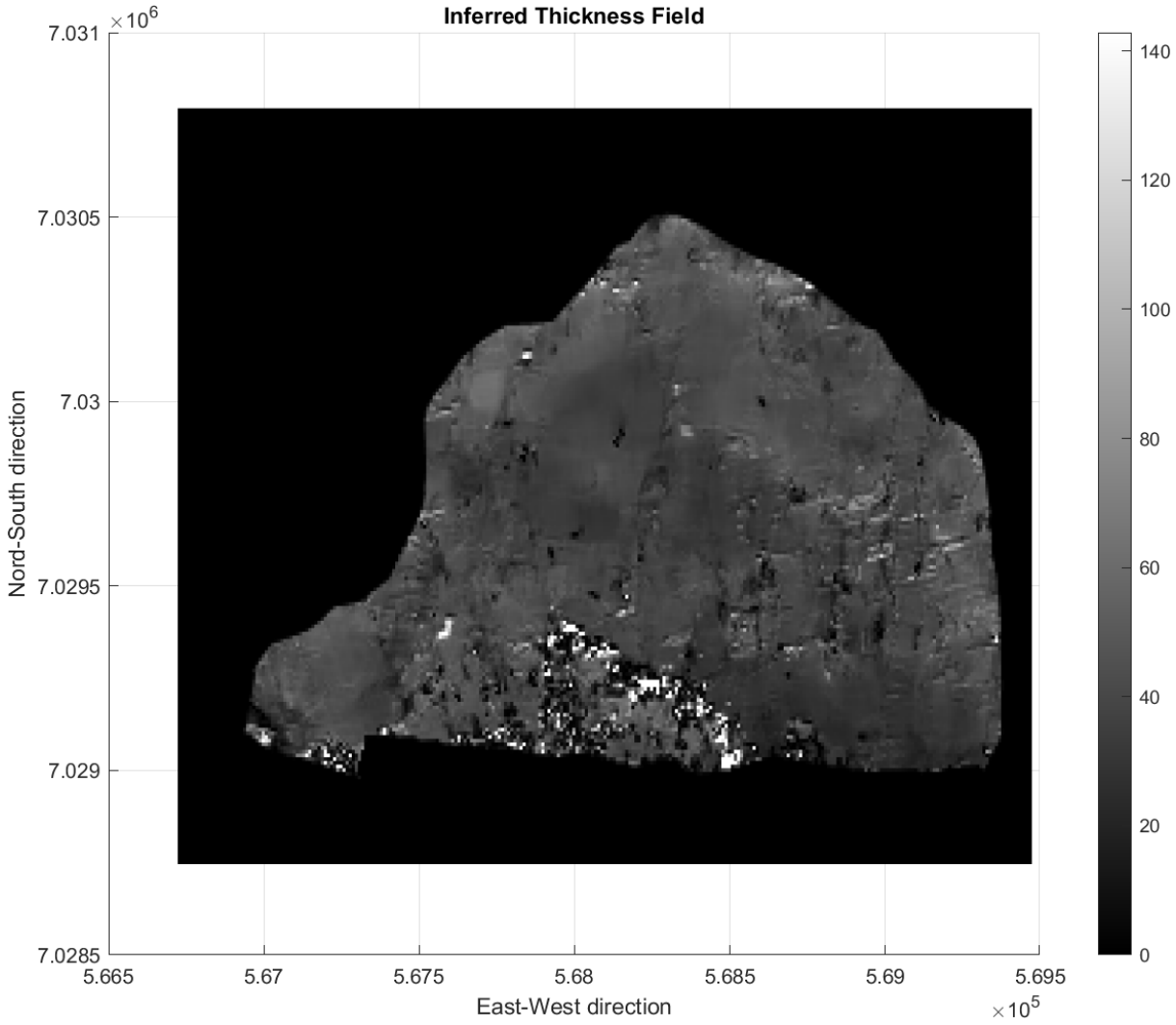


Figure 4.2: Numerical Experiment 1: inferred thickness field with meter as the unit of measurement.

In Figure 4.3 we plotted the best α satisfying the discrepancy principle which is constant on each cell. Most value of landslide are 0.01 which is different from 0 value outside the landslide, where we imposed the thickness to be the zero vector just as on the southern part corresponding to the glacier.

From the efficiency point of view, on a laptop Lenovo Legion 5 with 16GB of

RAM and processor AMD Ryzen 7 5800H the program took almost 2 days to complete the inversion.

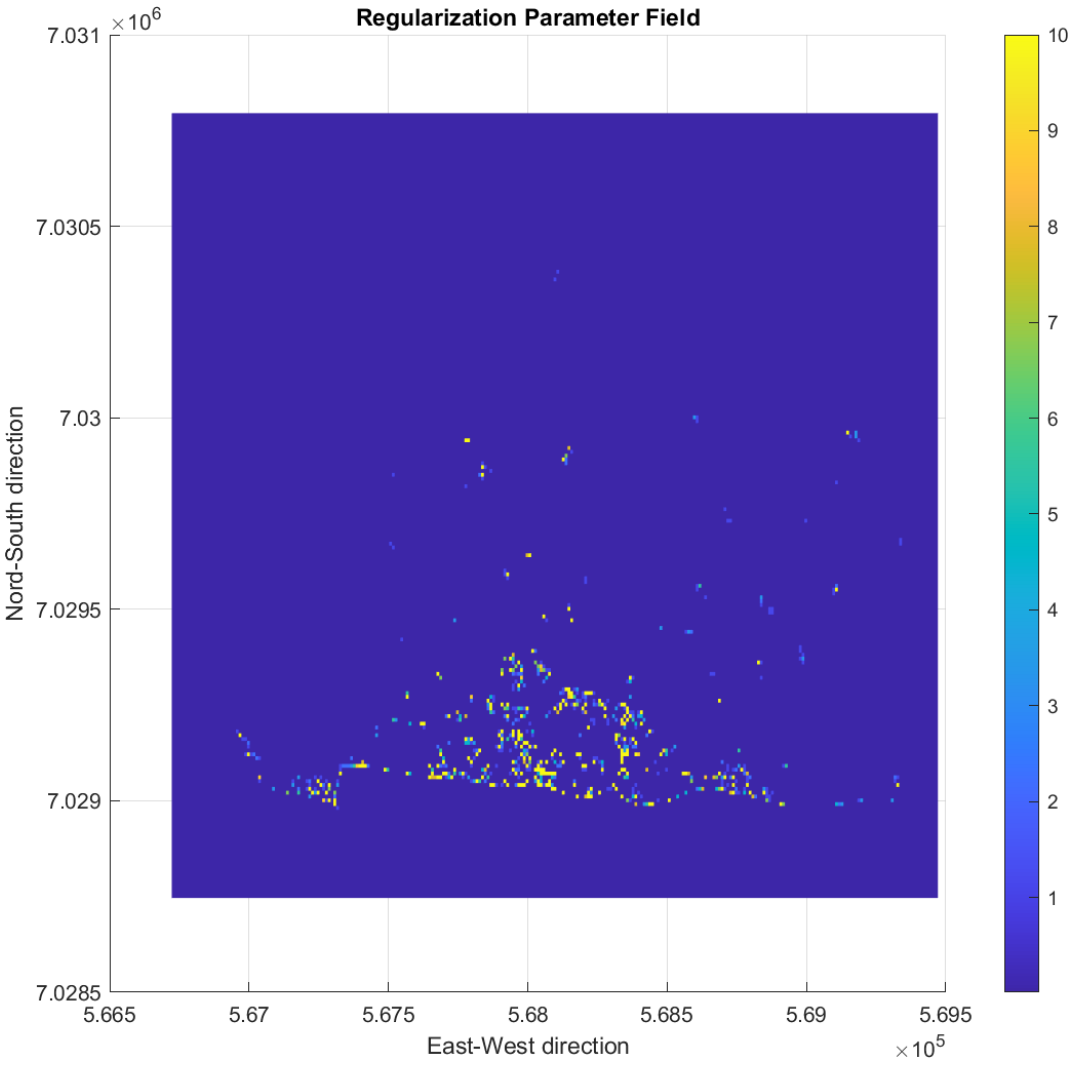


Figure 4.3: Numerical Experiment 1: Regularization parameter under square root.

4.2 Numerical Experiment 2

Since no information on the noise is known, we then implemented Algorithm 1 (Balancing Principle) in the inversion program such that no a priori bounds on the regularization parameter are set.

We chose $maxit = 10$ and $tol = 10^{-5}$ with a scaling parameter equal to $\frac{10}{N_{ls}}$ where N_{ls} is the number of data points on the active landslide. In Figure 4.4 we plotted the inferred mean thickness with the same threshold and the same scaling factor γ as in Experiment 1. Despite showing globally the same geometry, the result from balancing principle seems to offer more details than the one from approximate discrepancy principle.

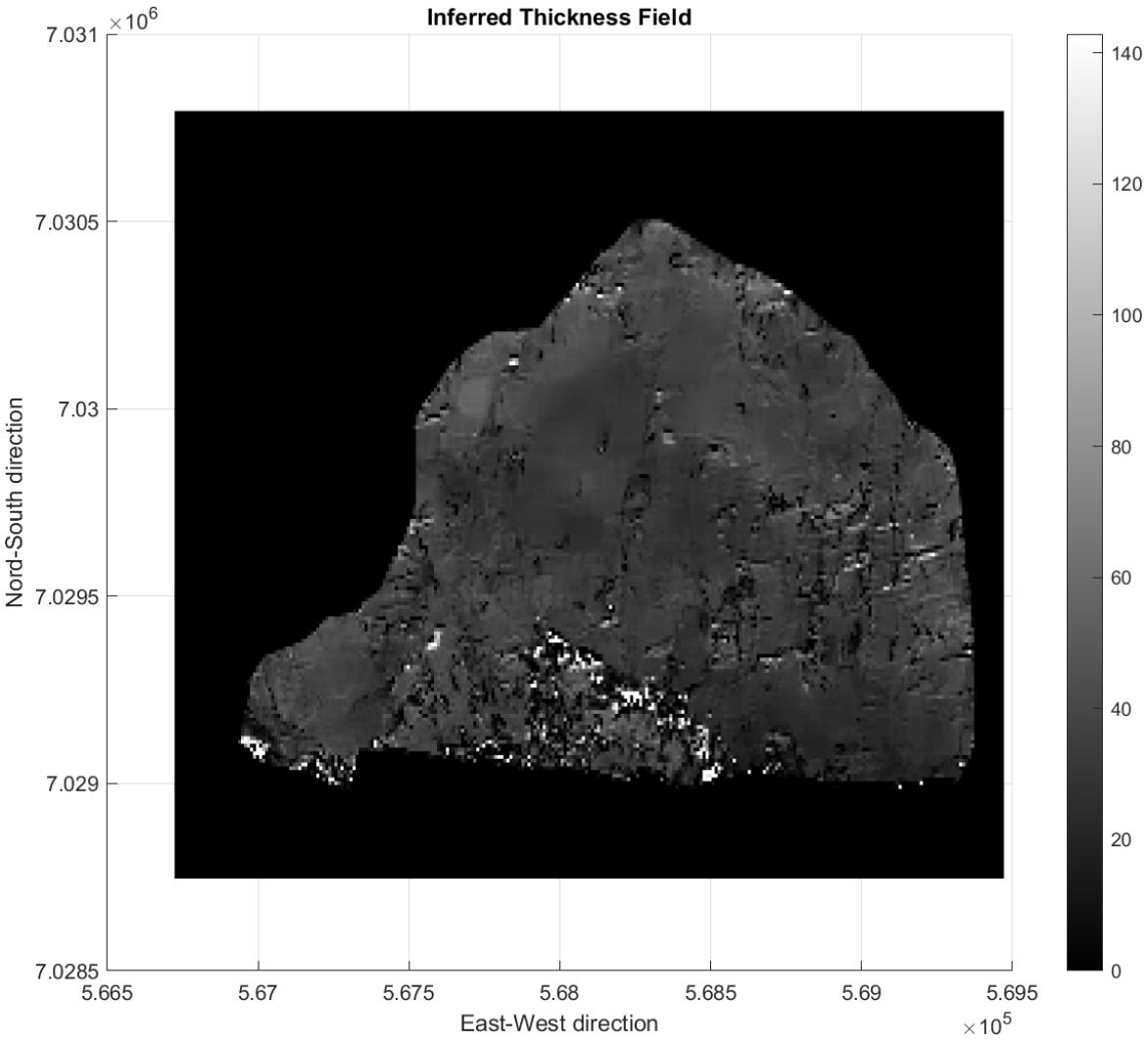


Figure 4.4: Numerical Experiment 2: inferred thickness field with meter as the unit of measurement.

This new principle provides directly λ in stead of α. From the graph of the regularization parameter (Figure 4.5) we can see peaks again on the southern part of the landslide, in particular extremely large values are reached ($\sim 10^{11}$). In order to compare with the previous result we plotted in Figure 4.5 the values of $\lambda < 10^2$, since we're interested in values between 10^{-4} and 10^2. The most evident fact is that balancing principle seems to be able to detect the contour of the landslide.

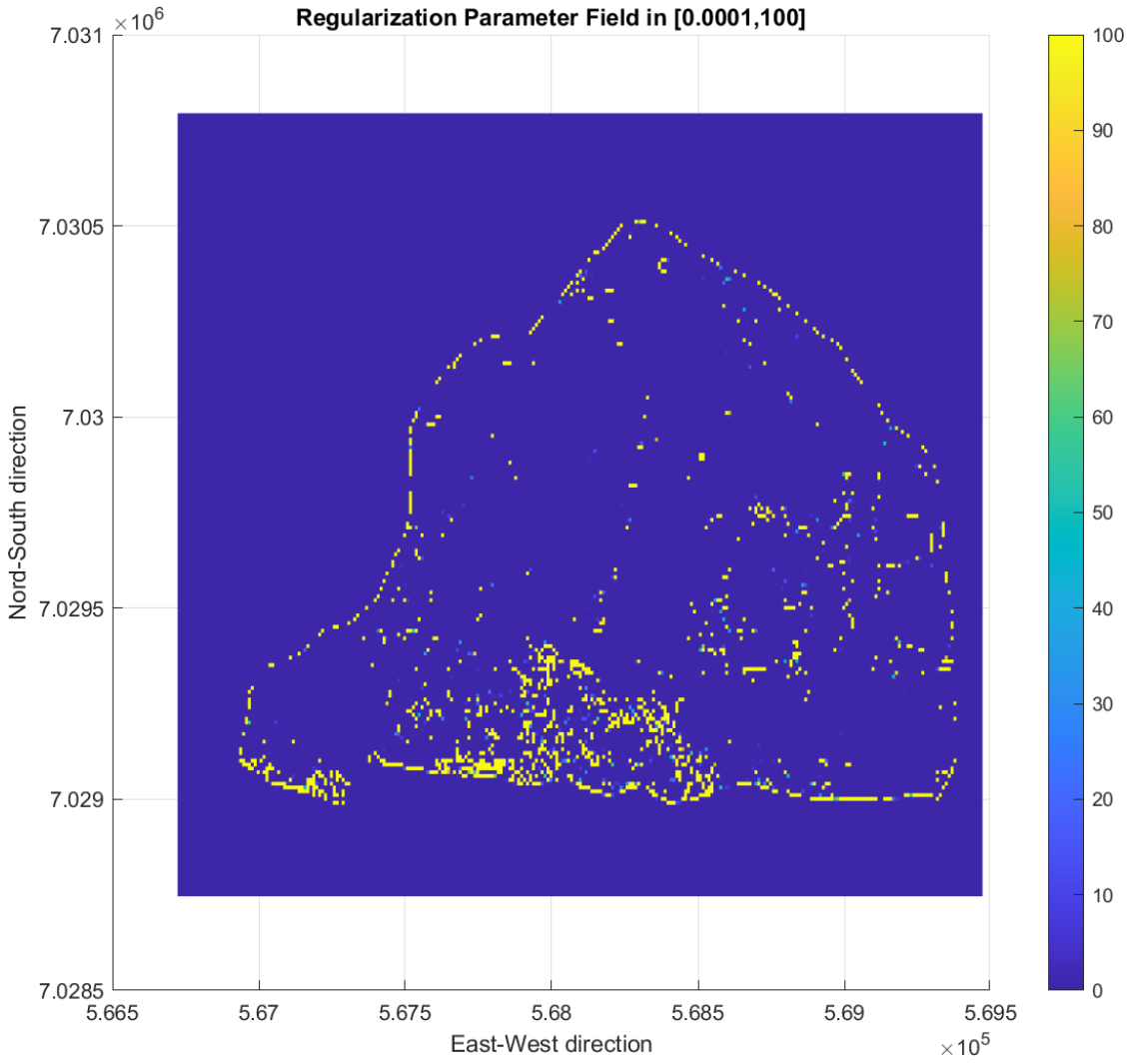


Figure 4.5: Numerical Experiment 2: regularization parameter less than 100.

In Figure 4.6 we plotted the number of iterations run by Balancing Principle cycle which has a constant value on each subgrid. On the same computer the inversion was completed in almost 1 day.

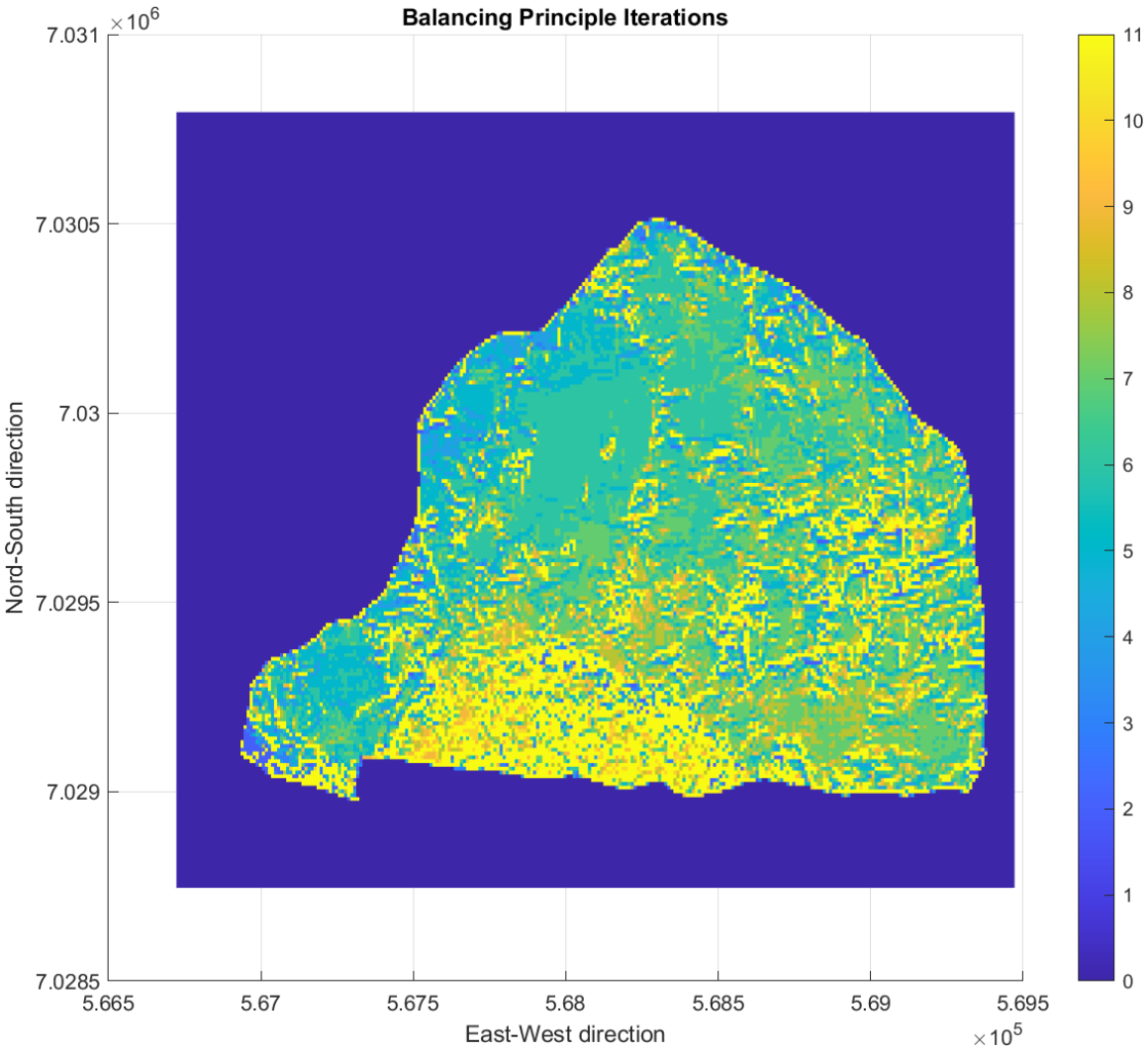


Figure 4.6: Numerical Experiment 2: number of iterations in Balancing Principle.

4.3 Numerical Experiment 3

Now instead of using the CVX program to solve the optimization subproblems, we can implement Algorithm 2 (Gradient Projection) with acceleration techniques provided in Algorithm 3 (SGP) by choosing the scaling matrix to be the identity matrix.

In particular we set the parameters as follows: $c_1 = 10^{-4}$, $\theta = 0.4$, $maxit = 70$, $tol = 10^{-5}$ and $[\alpha_{min}, \alpha_{max}] = [10^{-10}, 10^5]$. We took the zero vector as the initial guess. The tolerance used is a relative one, i.e. the program returns a result as soon as ration between the norm of the current gradient and the initial one is less than tol.

Essentially the result shown in Figure 4.7 is the same geometry found in Experiment 2 but with a scaling factor $\gamma = 0.028$.

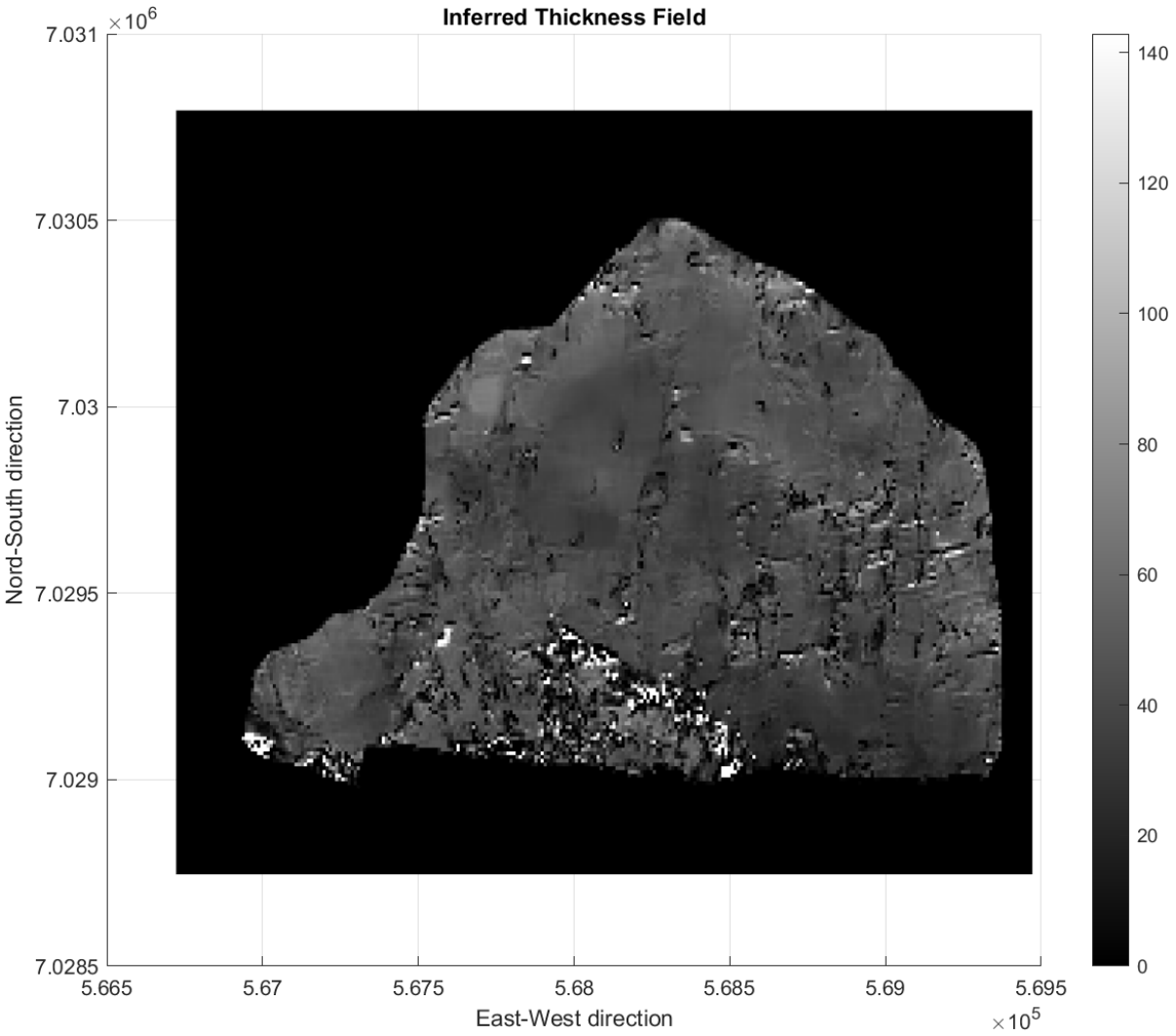


Figure 4.7: Numerical Experiment 3: inferred thickness field with meter as the unit of measurement.

Regarding the regularization parameter (Figure 4.8), the largest value is of order of magnitude $\sim 10^{12}$, whilst the patter of distribution of λ between 10^{-4} and 10^{2} is similar to the previous experiment. Also the number of iterations run by balancing principle is almost the same.

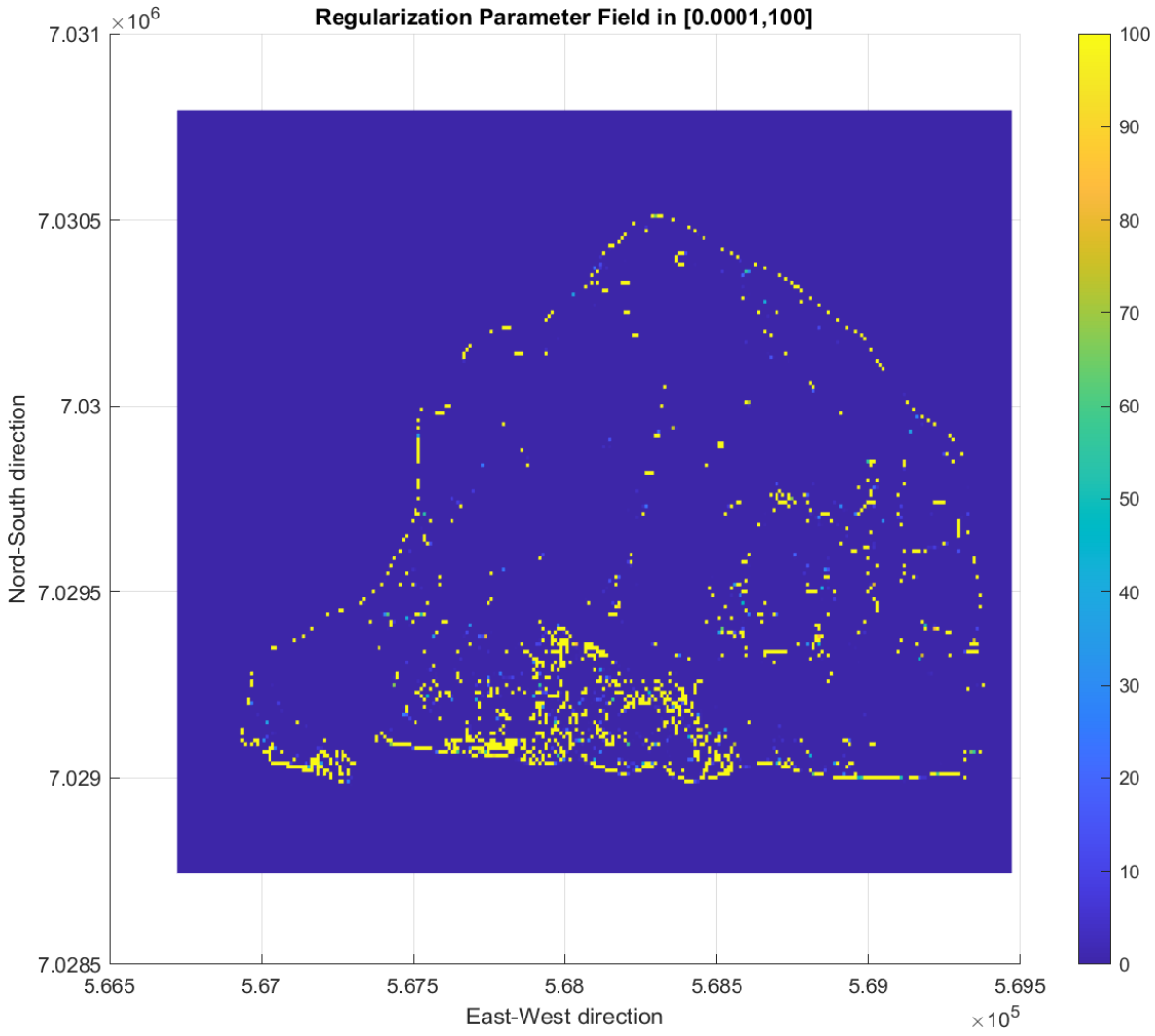


Figure 4.8: Numerical Experiment 3: regularization parameter less than 100.

In Figure 4.9 is shown the number of iterations run by Balancing Principle, and in Figure 4.10, the number of iterations run by gradient projection.

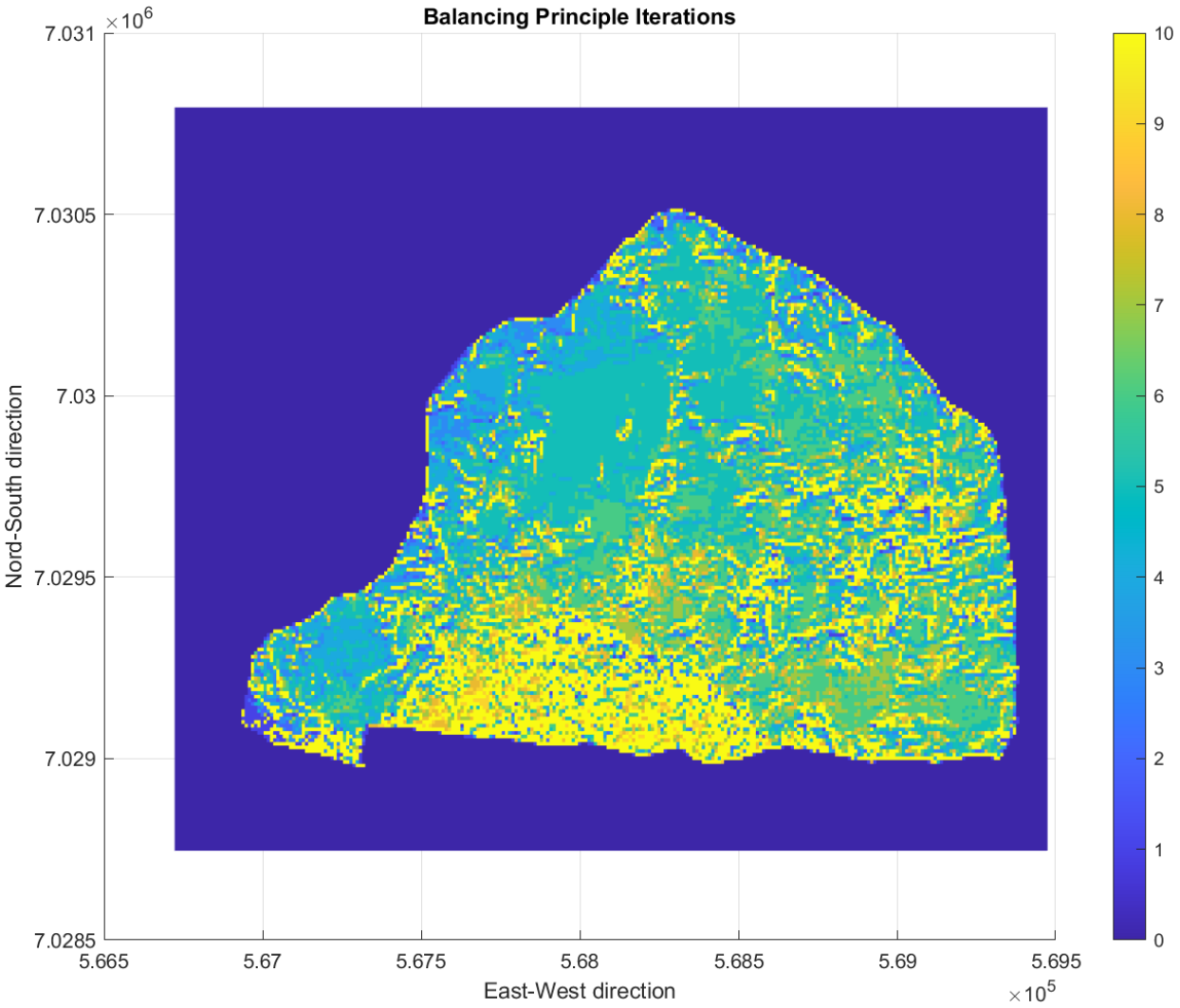


Figure 4.9: Numerical Experiment 3: number of iterations in Balancing Principle.

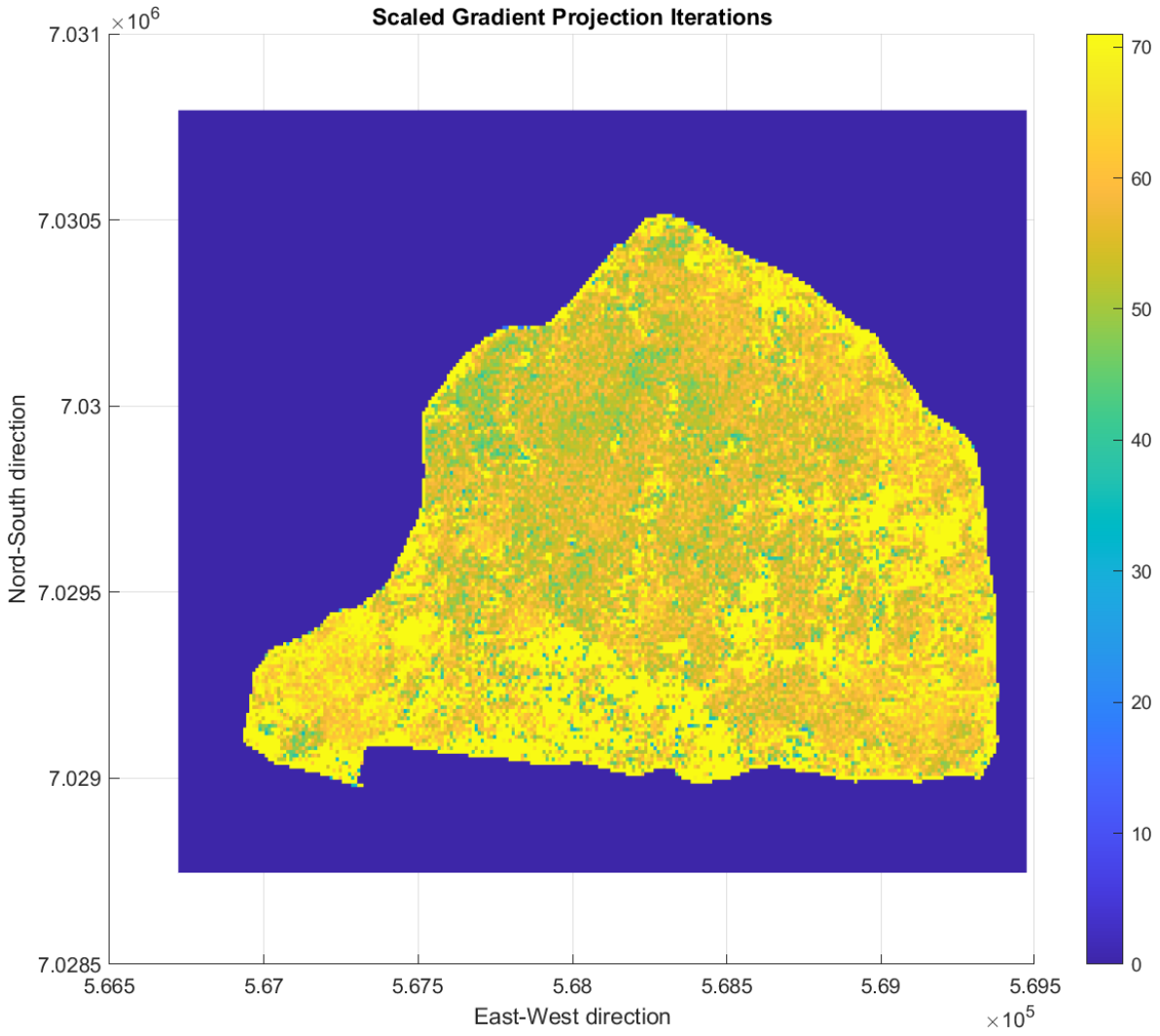


Figure 4.10: Numerical Experiment 3: number of iterations in SGP.

In Figure 4.11 and 4.12 we plotted the exiting ratio of norm of gradients and the exiting norm of gradient in logarithmic scale. We chose $maxit = 70$ to make sure that the optimization method provides a good enough result. The inversion program took 9 minutes and 10 seconds to give this result, a remarkable improvement given the similarity of the results.

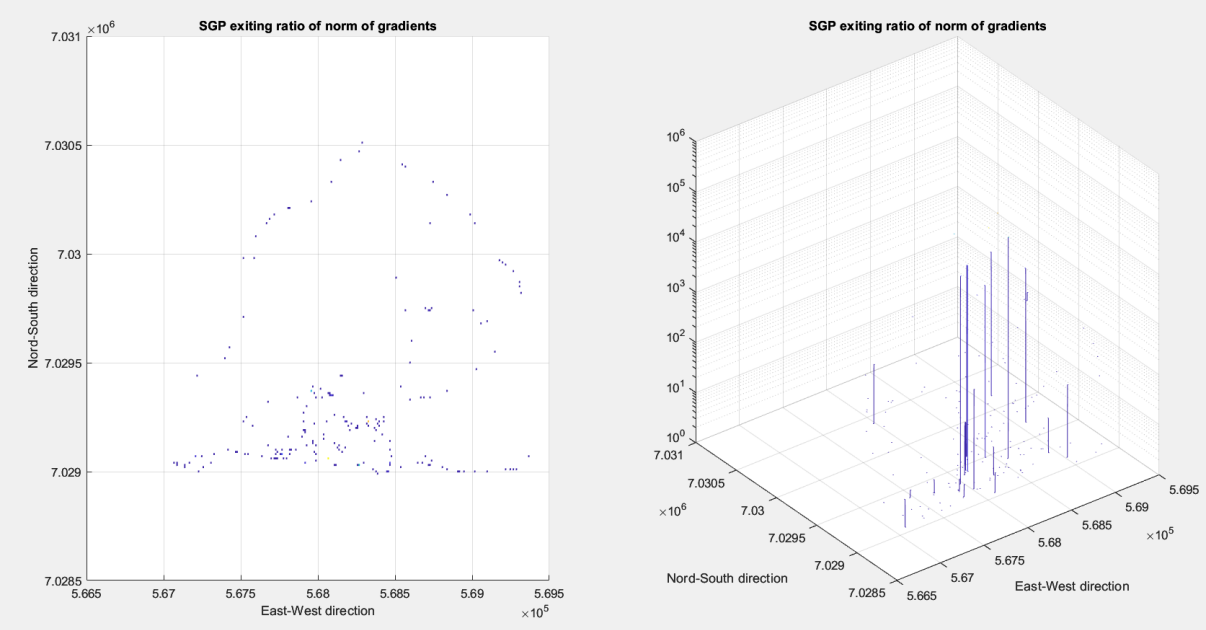


Figure 4.11: Numerical Experiment 3: ratio of norm of gradients in SGP.

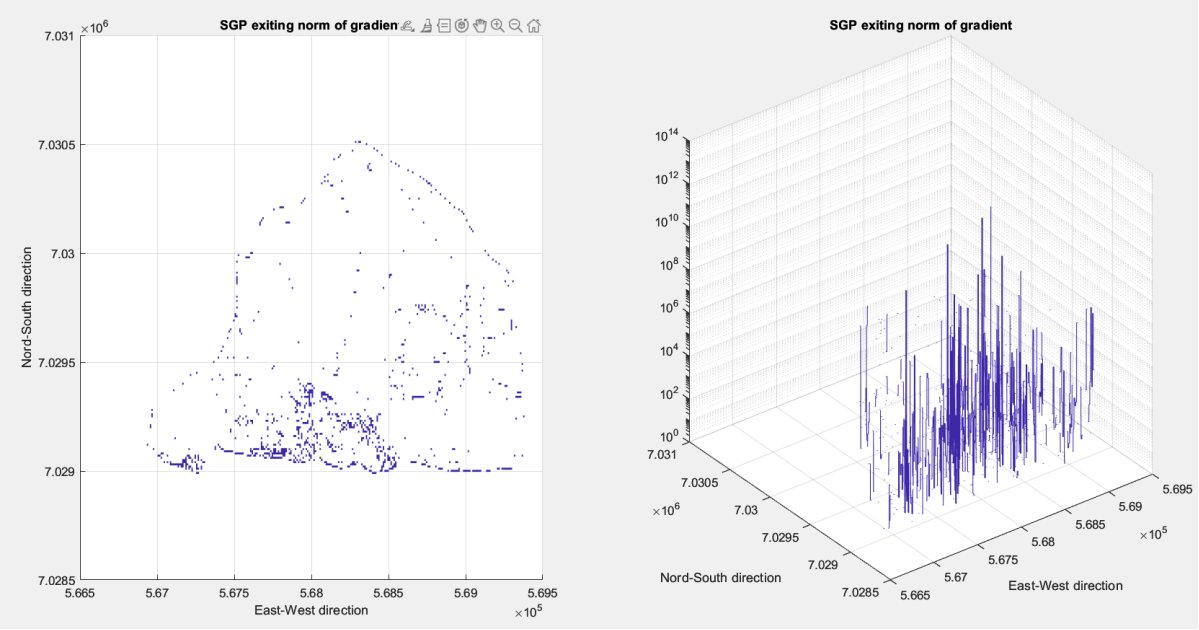


Figure 4.12: Numerical Experiment 3: norm of gradient in SGP.

4.4 Numerical Experiment 4

Finally we experimented the Algorithm 4 (CNPCG). We took the zero vector as the initial guess $x_0 = 0$. We set $\eta = 10^{-4}$, $maxit = 140$, we chose the tolerance on the norm of the gradient to be $tol = 10^{-5}$ and $\epsilon = 10^{-10}$ obtaining the results plotted in Figure 4.13. The scaling factor of the inferred thickness is chosen as $\gamma = 0.021$.

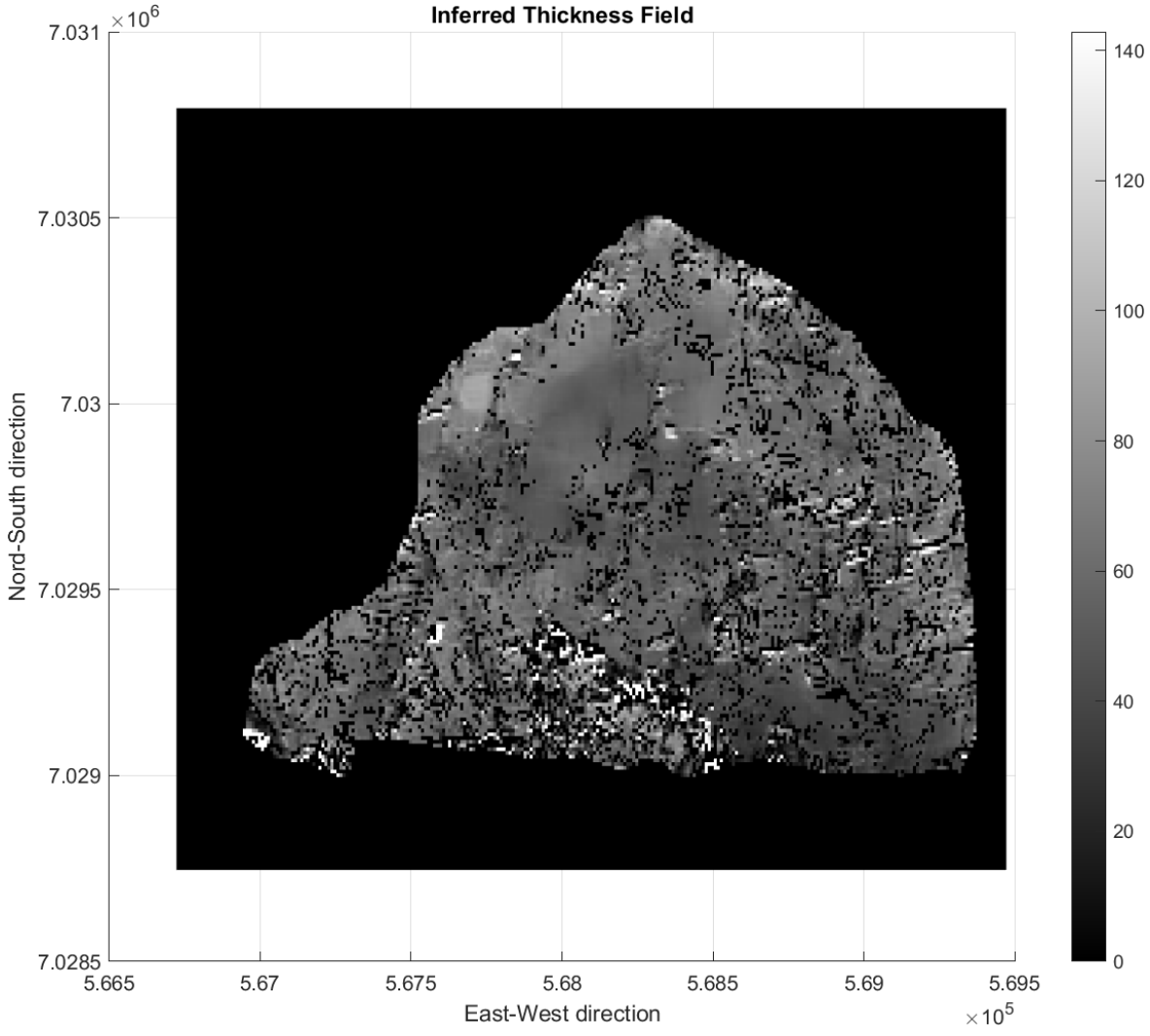


Figure 4.13: Numerical Experiment 4: inferred thickness field with meter as the unit of measurement.

We also decided to double the number of iterations in the balancing principle cycle, $maxit = 20$. It's been observed a generally higher values of regularization parameters as shown in Figure 4.14.

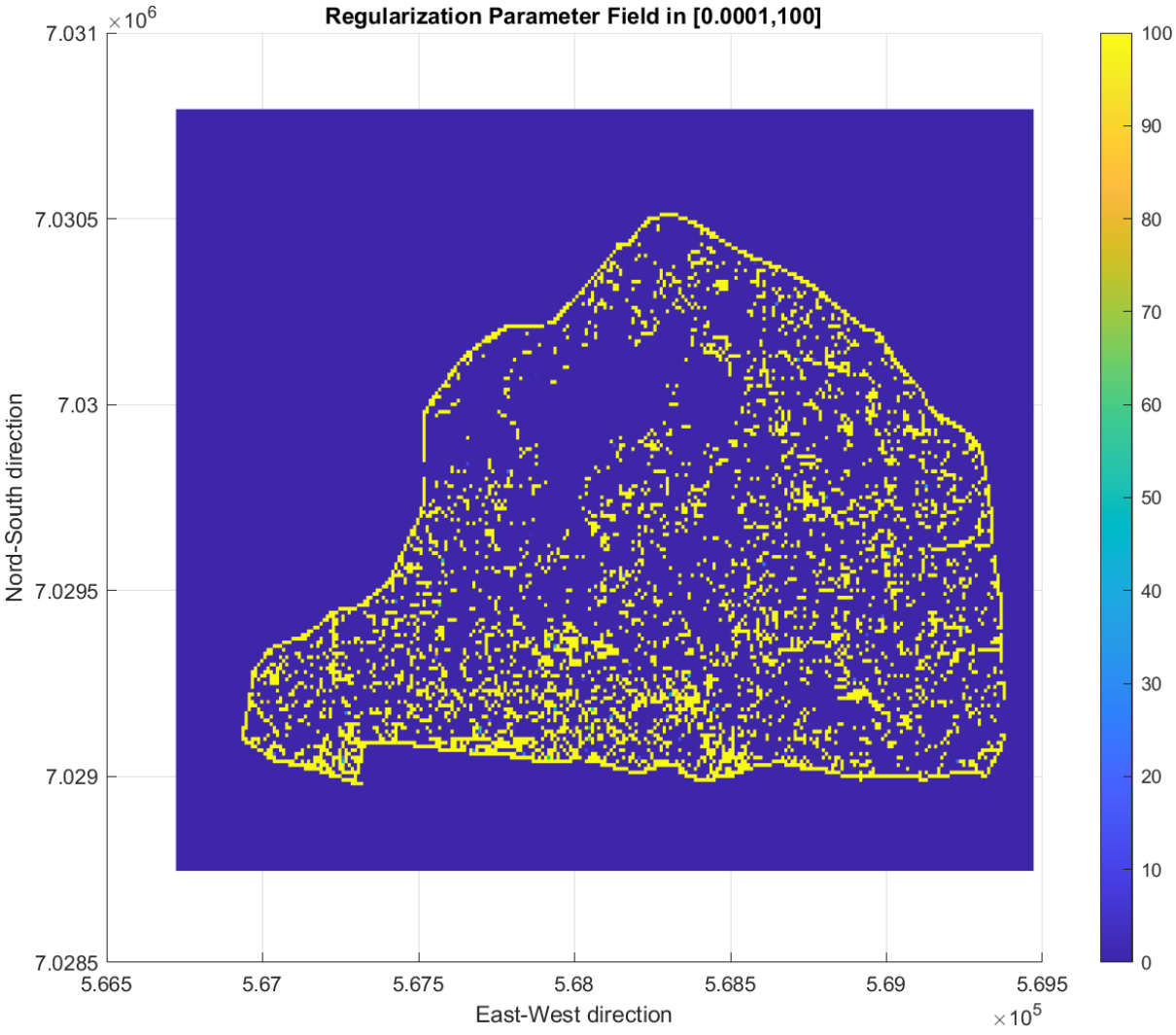


Figure 4.14: Numerical Experiment 4: regularization parameter less than 100.

In Figure 4.15 is shown the number of iterations run by Balancing Principle, and in Figure 4.16, the number of iterations run by Newton projection.

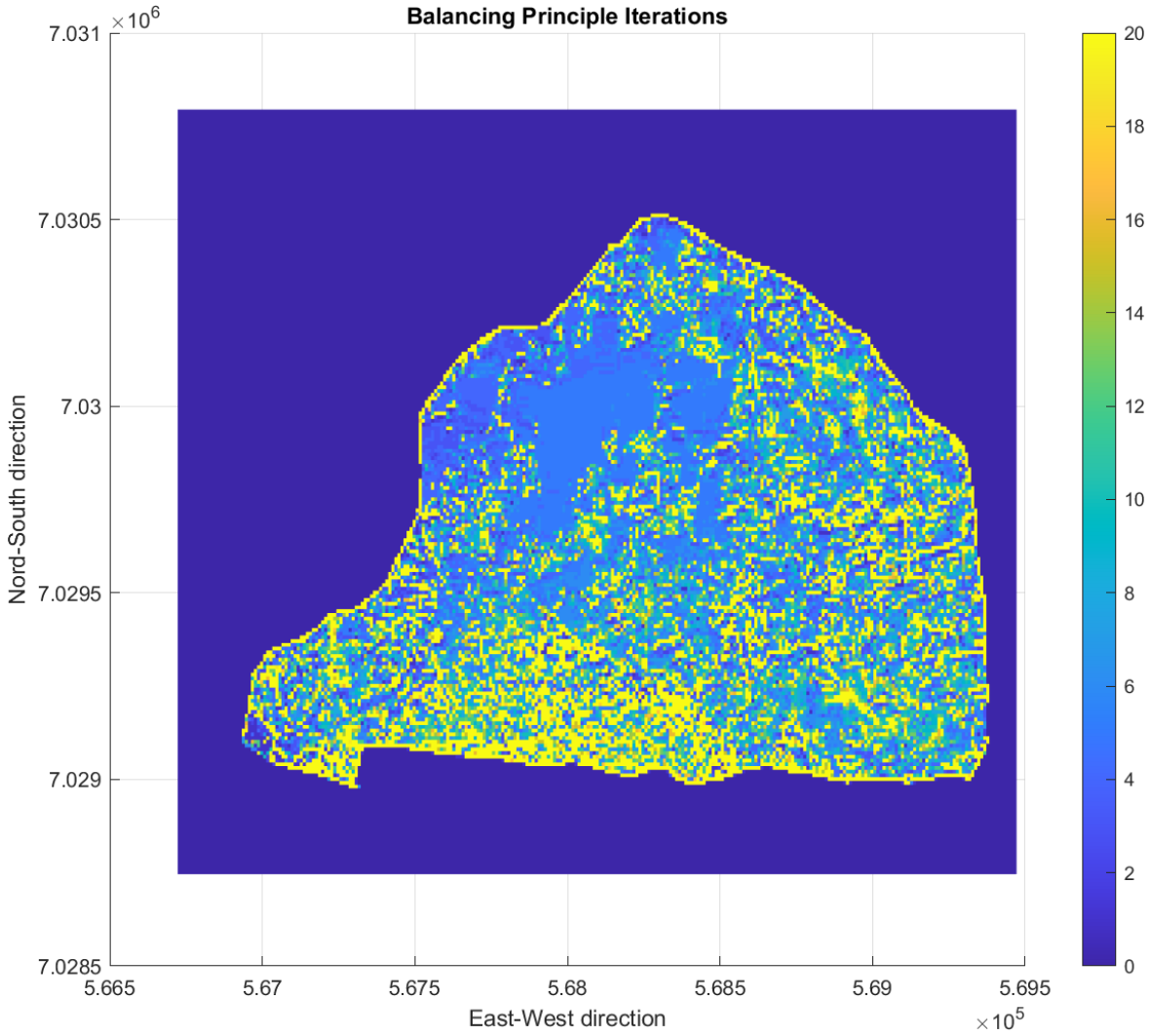


Figure 4.15: Numerical Experiment 4: number of iterations in Balancing Principle.

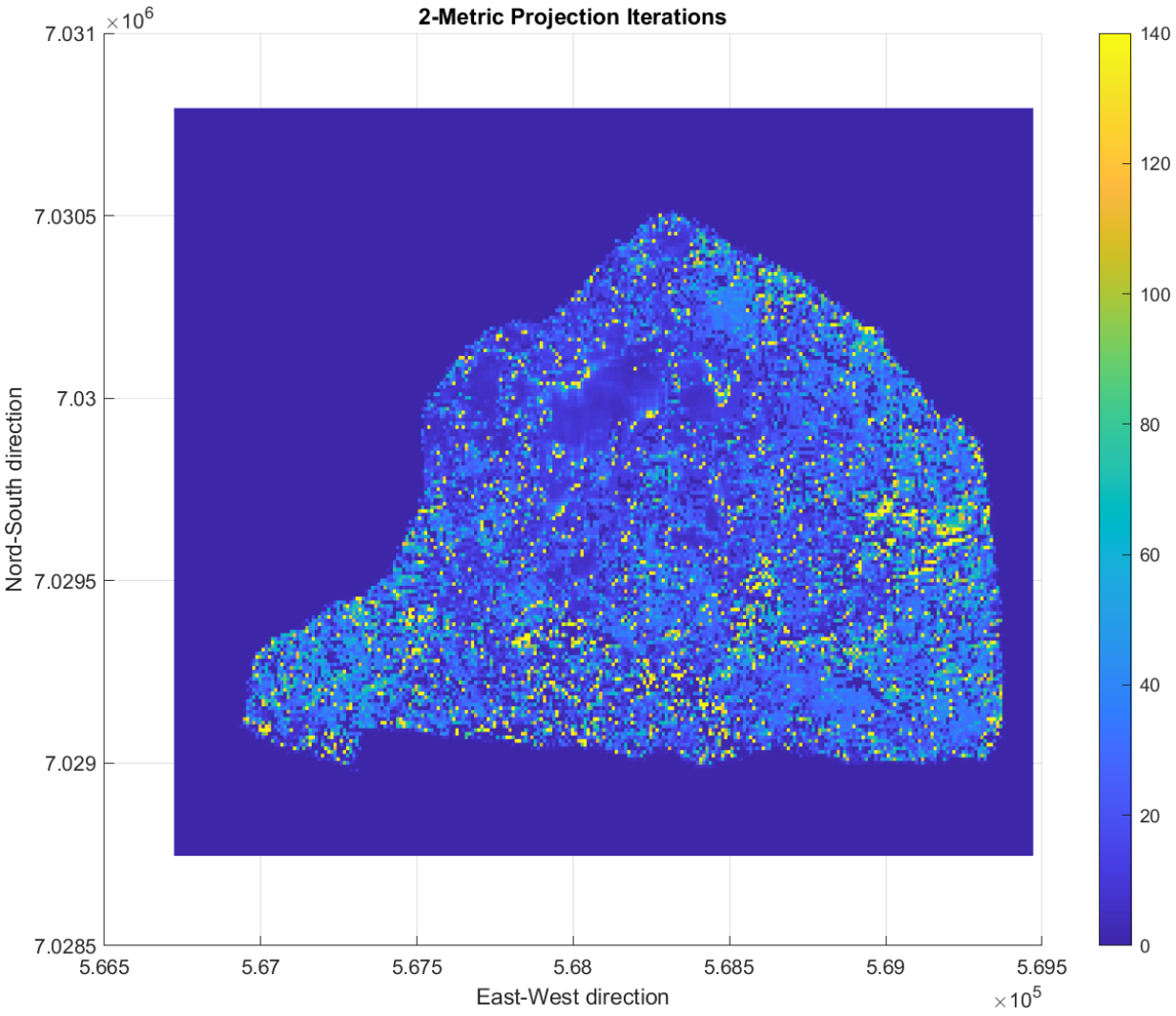


Figure 4.16: Numerical Experiment 4: number of iterations in CNPCG.

Finally in Figure 4.17 we plotted the norm of the projected gradient exiting the Newton projection cycle in logarithmic scale. The program took 1 hour and 10 minutes to complete the inversion.

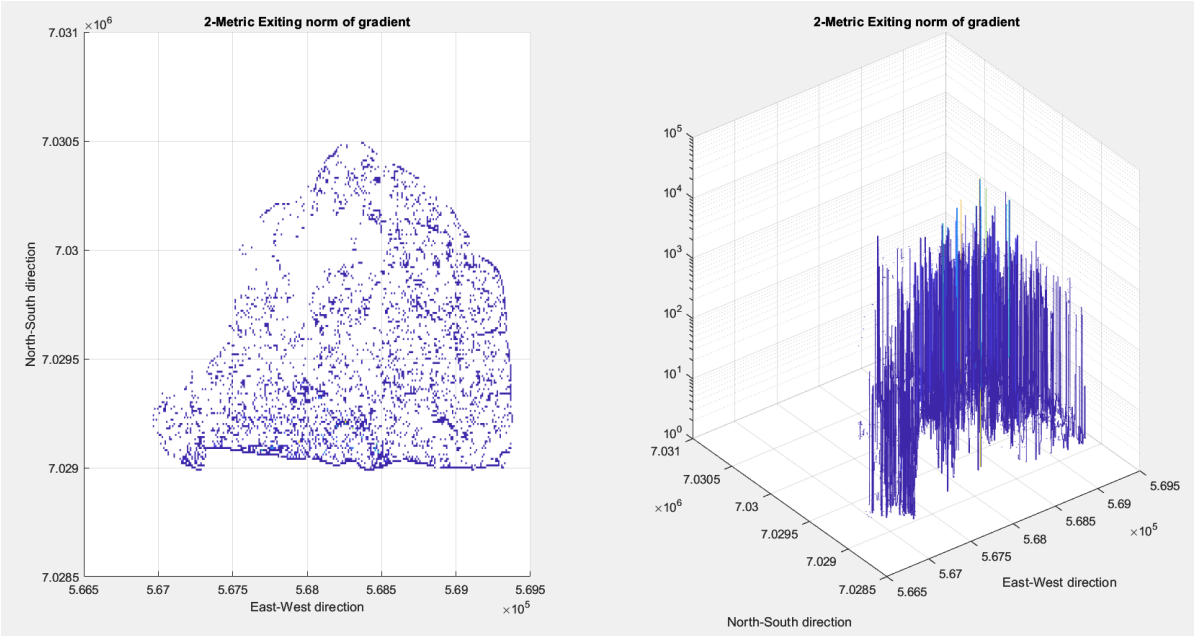


Figure 4.17: Numerical Experiment 4: norm of gradient in CNPCG.

4.5 Discussion

The overall geometries found by all four implementations are quite similar, but the inversion method with balancing principle and SGP is the most efficient one.

Tikhonov regularizer chosen by Booth et al. (2013) has smoothened to some degree the results. Further improvements might be achived by choosing a L1-norm based regularizer such as the total variation regularizer (Di Serafino et al., 2020) or by considering multi-penalties as done in [36] by Bortolotti et al. (2021).

Chapter 5

Interpretation

In this chapter we review some results on Fels landslide obtained by engineering geologists and Earth scientists.

In particular, first we review a classical method used to infer the thickness of landslides, then we shortly present the latest results published by Donati et al. (2021) and by Rabus et al. (2022).

5.1 Vector Inclination Method

We've seen in Theorem 2 that the proposed inversion method provides a scaled version of the underlying thickness. If even a single direct measurement of local thickness can be made, for example, in a borehole or with shallow geophysical techniques, that measurement can be used to determine the scaling factor by minimizing the misfit between the observed and inversion-predicted thickness.

Unfortunately no direct measurements are available for Fels landslide. Therefore we chose a scaling factor for each experiment such that the resulting thickness values belong to the range [0,140]. This choice is justified by the result obtained with a classical estimation method we used to infer the thickness but only along a profile of the landslide. The method is known as the

Vector Inclination Method (VIM) [30].

The VIM starts with the assumption that a point on the surface of the landslide will move in a direction that is parallel to the slope of the sliding surface beneath, provided that the mass moves as a rigid body. Another assumption is that a single sliding surface exists. Also it's required that a displacement monitoring must be implemented and measurement points (MPs), taken along a cross-section of choice, must be available. Then, the methods can by applied as follows:

1. Draw the cross-section of the landslide intersecting the MPs and mark with arrows the vectors of movement $\bar{V}_i$ measured at each MP. The first vector should be representative of the movement close to the main scarp.

2. Draw the normals to each vector. Then find the intersection points O_i between two consecutive normals and draw the bisection lines between two consecutive normals.

3. The first point of the sliding surface P_1 is identified as the first MP. The second point P_2 is found as the intersection of the line generate by $\bar{V}_1$ and the first bisection line. The third point P_3 is found as the intersection of the line generated by $\bar{V}_2$ applied to P_2 and the second bisection line, so on and so forth.

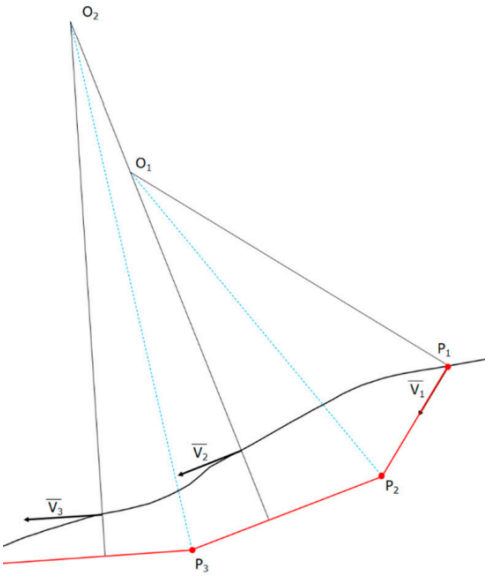


Figure 5.1: Vector Inclination Method procedures. Picture from [30].

Dr. Davide Donati and his research team traced various profiles across the Fels slope. They sampled the depth of the reconstructed surface and then interpolated the points to produce a surface shown in figure 6.3.

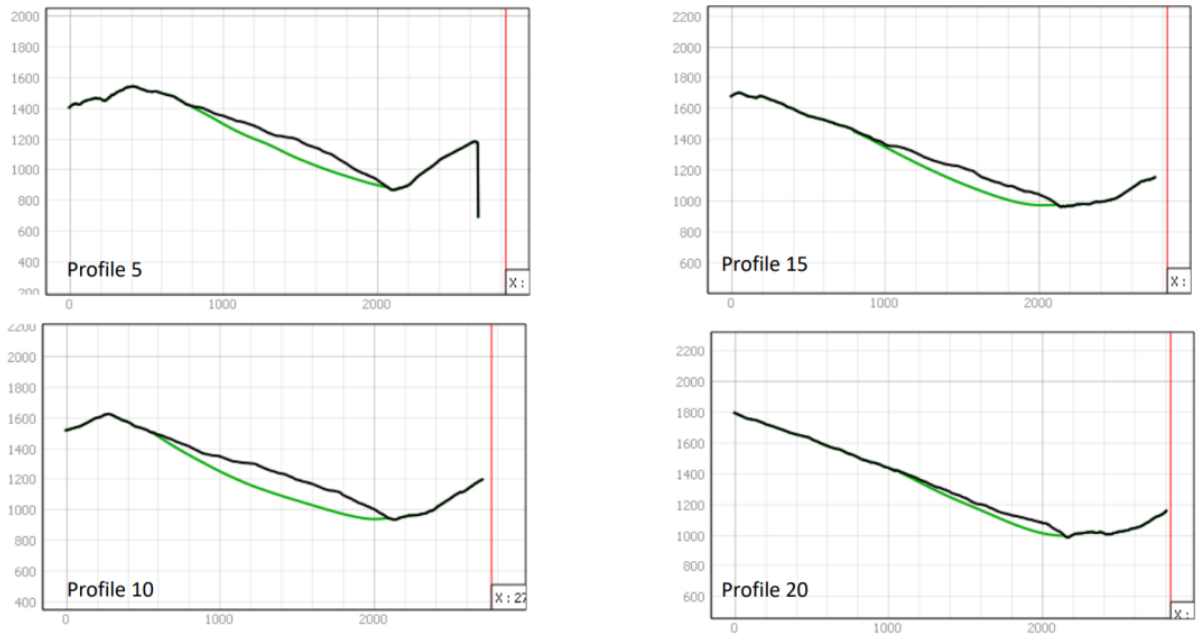


Figure 5.2: Examples of profiles analyzed. Courtesy of Dr. Davide Donati.

5.2 Profiles Analysis

Donati et al. (2021) derived the trend, plunge and magnitude of displacement from data presented in Section 1.2 using the Geographical Information System (GIS) software ArcGIS 10.6. Then they extracted some profiles parallel to displacement direction and compared them with elevation data. The combined analysis of displacement, plunge and elevation profiles facilitated identification of geomorphic features produced by slope deformation such as uphill- and downhill-facing scarps, grabens and slope breaks.

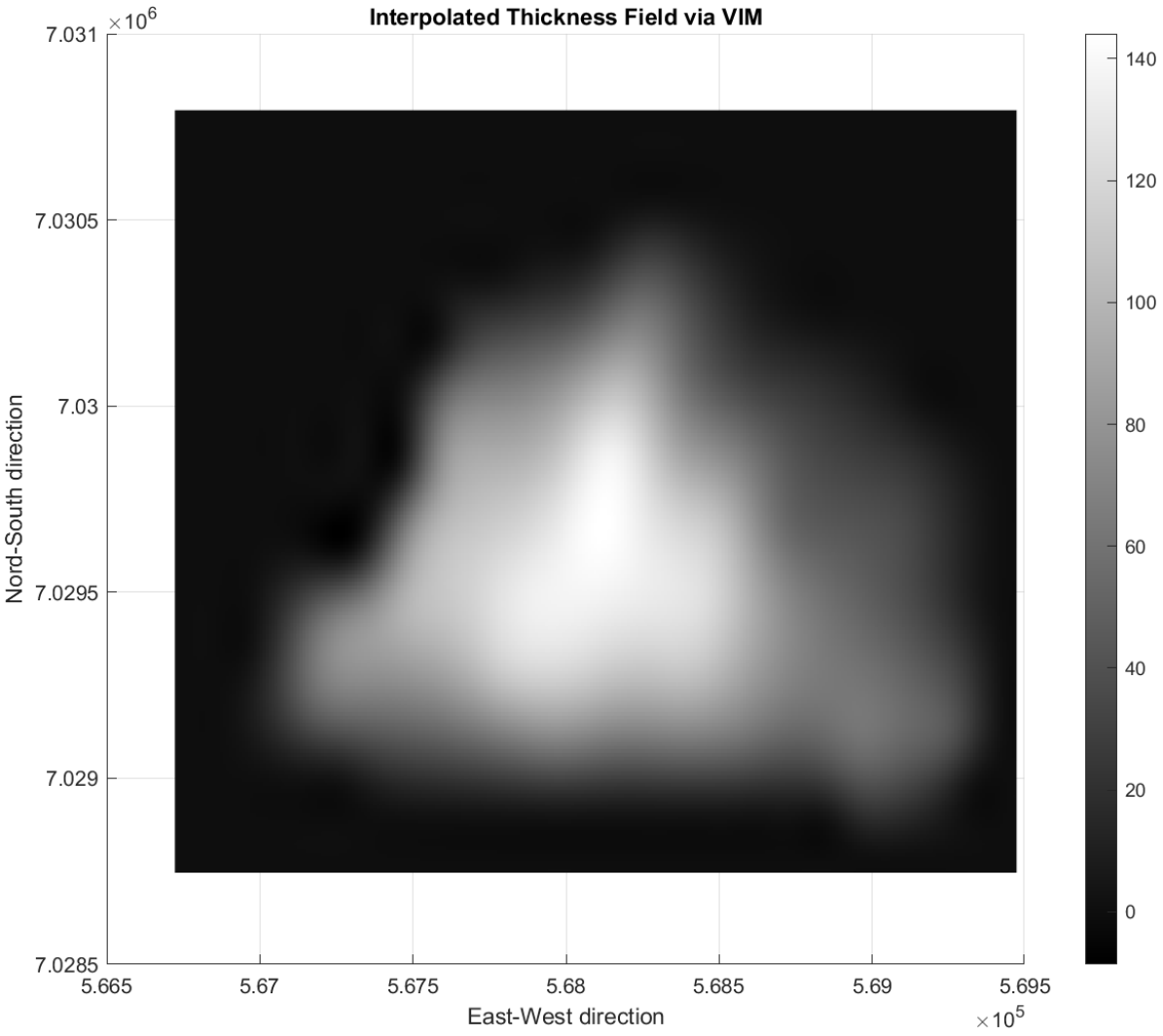


Figure 5.3: Interpolated Thickness Field via VIM with meter as the unit of measurement. Courtesy of Dr. Davide Donati.

One of such profile is the AA' shown in Figure 5.7 (c). The displacement profiles display a constant magnitude in the upper slope, suggesting that the landslide body is deforming as a rigid body. A progressive downslope increase in displacement magnitude begins in the central part of the slope and continues downward to the boundary with the fast-moving slope toe. Low-order undulations are evident within this zone, which locally cause the cumulative displacement to decrease downslope (Figure 6.7 (a)). These

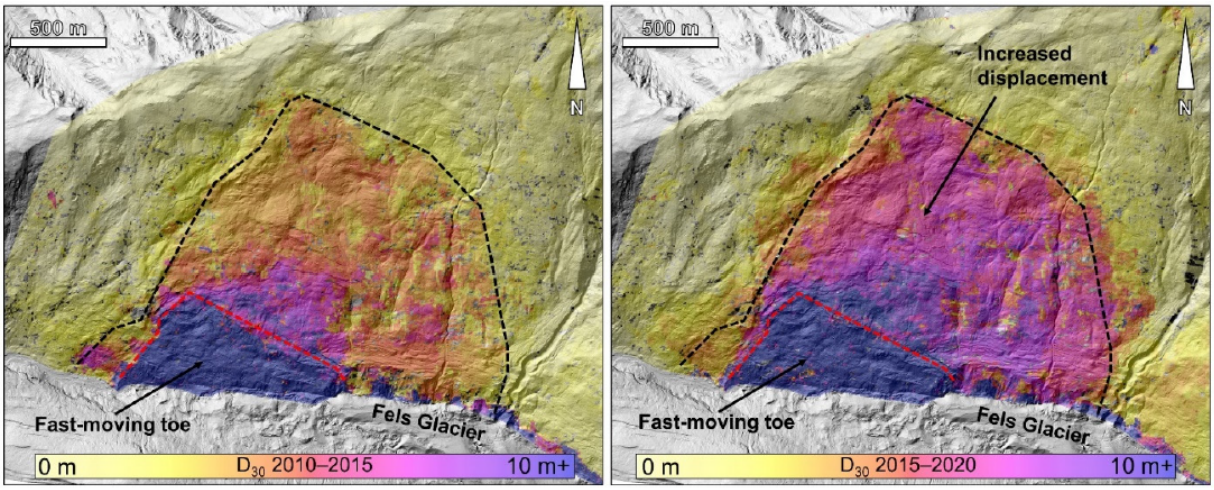


Figure 5.4: Displacement magnitude field in 2010-2015 windows (a) and 2015-2020 window (b). Picture from [1].

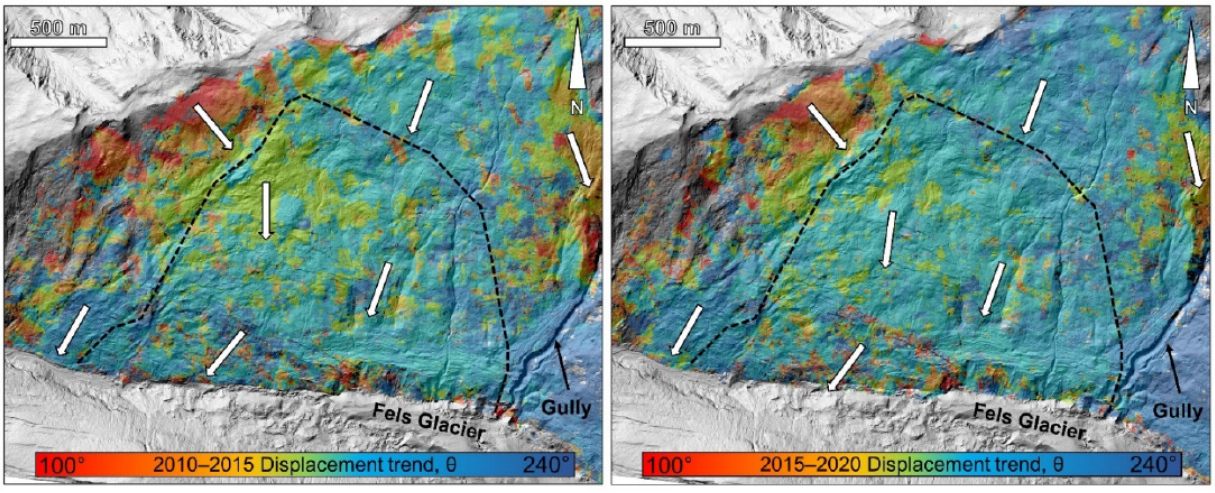
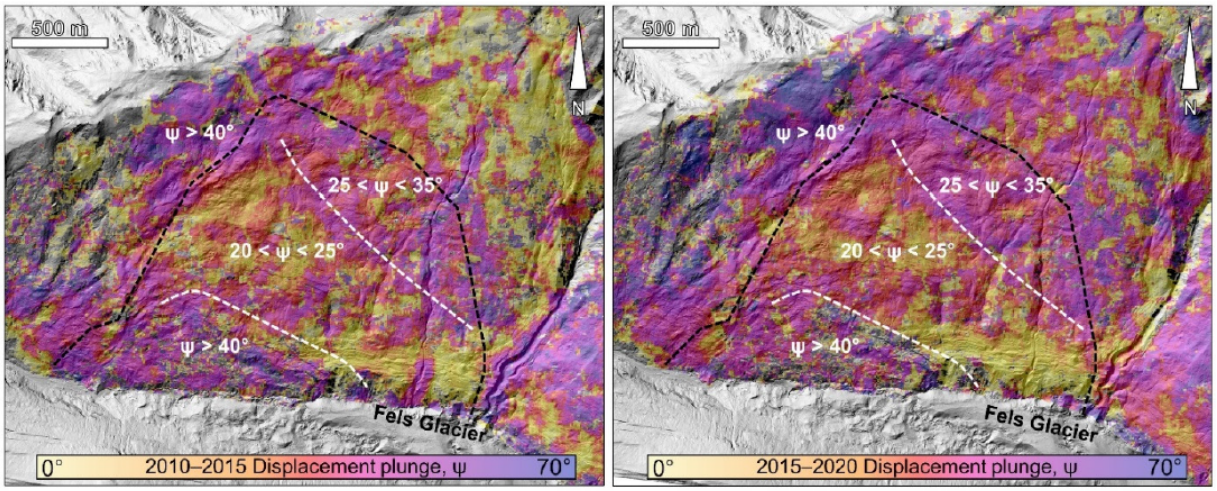


Figure 5.5: Trend field in 2010-2015 windows (a) and 2015-2020 window (b). Picture from [1]

undulations may be caused by shallow secondary instability within the colluvial deposit blanketing the slope (Figure 6.7 (c)). The transition between the uniformly displacing part of the slope and the part with an increase in



(a) (b)

Figure 5.6: Plunge field in 2010-2015 windows (a) and 2015-2020 window (b). Picture from [1].

displacement roughly coincides with a break in the elevation profile where the slope steepens. A slope break also marks the upper boundary of the fast-moving toe. This complex pattern in displacement magnitude is in accord with the deformation plunge profile (Figure 6.7 (b)). Undulations in amplitude are observed along the profile. These undulations are superimposed on the larger scale trend of the profile, which shows an overall decrease in plunge in the upper slope, between the headscarp and the slope peak at 1300 above sea level, followed by an increase in plunge in the central and lower slope. The area of increased plunge roughly corresponds to the increase in displacement magnitude in the central part of the slope. The decrease in part of the profile lower on the slope corresponds to the fast-moving toe, where significant undulations in plunge occur.

5.3 Structural Lineaments Comparison

They then compared the magnitude and distribution of the surface displacements with the location, distribution and orientation lineaments iden-

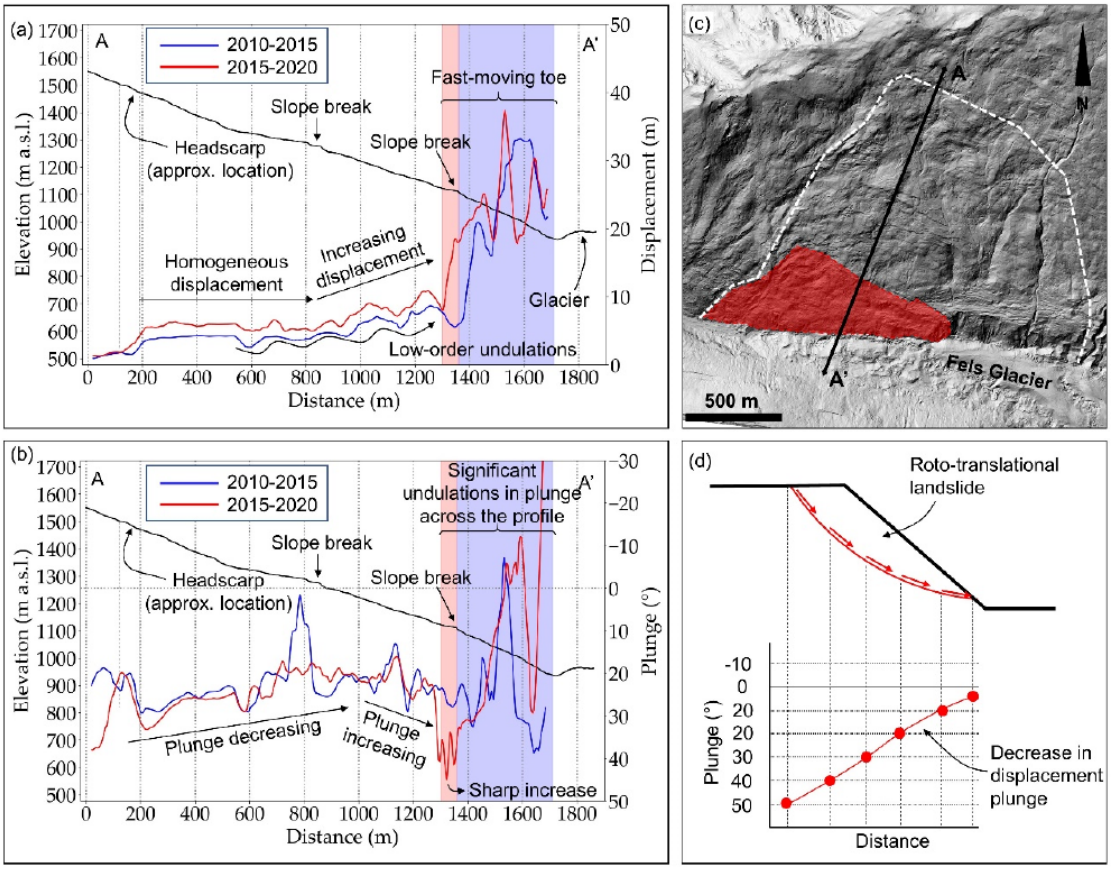


Figure 5.7: AA' (c) profile analysis. Picture from [1].

tified and mapped from the Airborne Laser Scanning (ALS) dataset. In particular they first identified three different lineament trends, I, II and III at Fels landslide using the ASL dataset. The lineament of the first type is NE-SW oriented, of the second type is NNW-SSE oriented and of the third type is NW-SE oriented.

Comparison of the lineament and displacement magnitude maps show significant correlation between the datasets. The boundaries of the landslide area can be roughly approximated by straight lines, with orientation similar to the identified lineament trends. The western boundary of Fels landslide is parallel to lineament trend I, and its location corresponds to the part of slope where the same trend has a high L_{21}, **lineament intensity** defined as

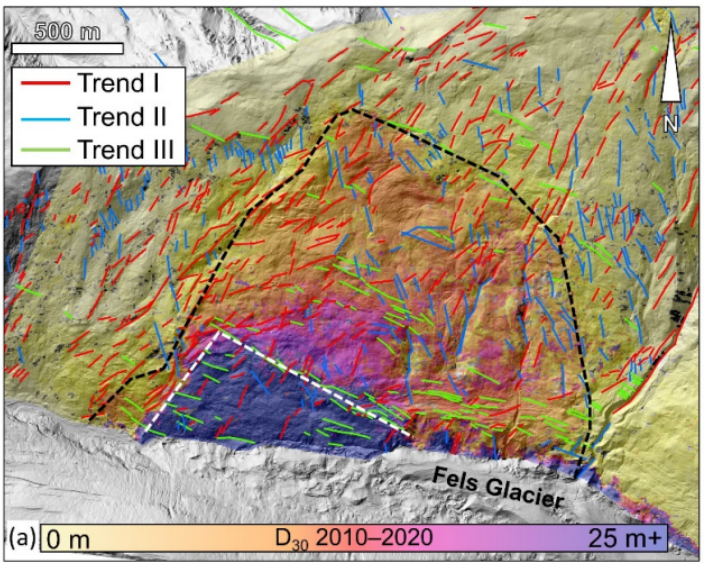


Figure 5.8: The three types of Lineament trends. Picture from [1].

the ratio of the total length of lineaments to the sampling area and measured in $[\frac{1}{m}]$. The orientation of the eastern and upper boundaries of the landslide area are sub-parallel with trend II and trend III lineaments, respectively.

The fast-moving toe appears to be outlined by scarps parallel to trend I on the west side and trend III along the rear boundary. From these observations they concluded that the landslide is structurally controlled and the surface deformation is mainly related to displacements originating at depth within the bedrock, rather than within the colluvial blanket draping the slope.

Based on reconstruction in Section 6.1 and the results from previous two sections, they proposed a subdivision of the Fels landslide into three different domains: upper, central and lower.

The upper domain (UD) is a slowly moving block with displacement rates of up to $1\frac{m}{year}$. Here, slope-parallel deformation vectors suggest that the displacement is occurring along a sub-planar sliding surface, likely controlled by foliation, that connects with the ground surface in the upper part of the domain. The head scarp does not correlate with any prominent geomorphic

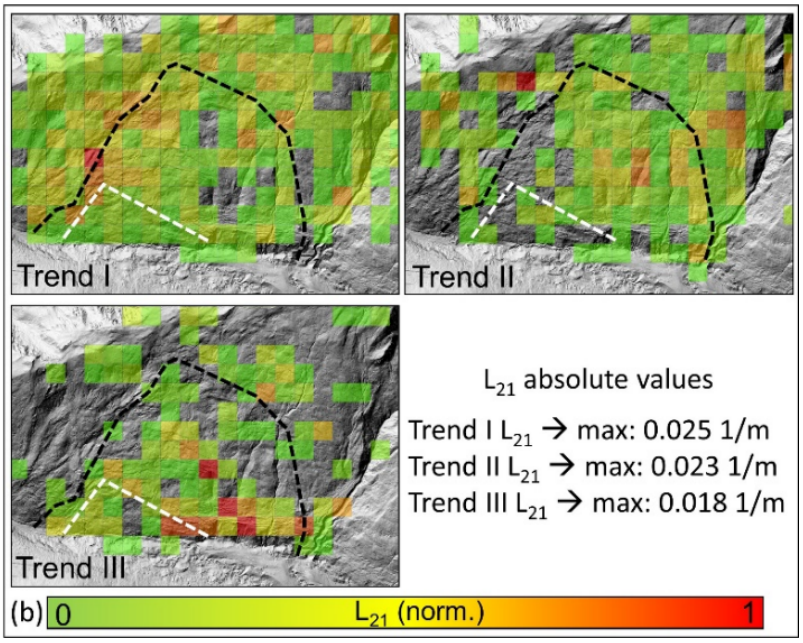
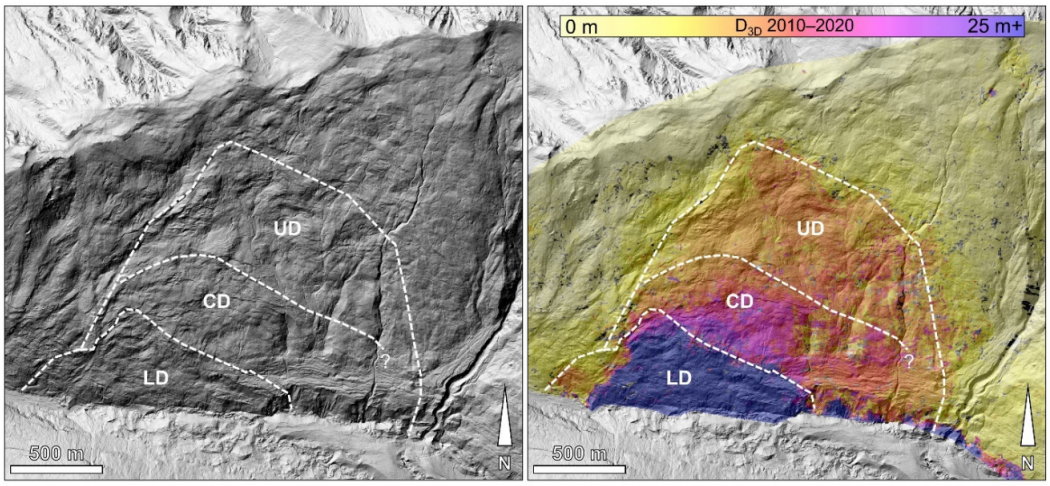


Figure 5.9: The lineament intensity of the three types of trends. Picture from [1].



(a) (b)

Figure 5.10: Three domains proposed by Donati et al. (2021) with picture from [1].

feature, suggesting that displacements noted in the upper slope are relatively recent, and colluvial material may be smoothing the ground surface. The style of deformation of the UD remained unchanged between 2010 and 2020.

The lower domain (LD) is the fast-moving toe that appears to be displacing by a rotational or pseudo-rotational mechanism, with displacement rates up to 5-8 $\frac{m}{year}$ over the 2010-2020 period.

The central domain (CD) appears to be transitional in nature between the lower, fast-moving domain and the upper, slow-moving domain. It's characterized by displacement rates ranging from 1 to 2 $\frac{m}{year}$, with a progressive increase in the downslope direction. A progressive increase in plunge is also noted across the domain, with highest values observed toward the fast-moving toe, indicating steepening and lengthening of the displacement vectors. The observed increase in plunge is likely not related to a steepening of the sliding surface, but rather to a local progressive change in failure mechanism from planar sliding to pseudo-rotational mechanism. The transition between planar sliding and slumping is also evident from the changes in the displacement plunge profile at the boundary between the CD and the fast-moving toe. The increase in displacement plunge shows that the volume upslope of the fast-moving toe was displacing through a planar sliding mechanism between 2010 and 2015, before transitioning to pseudo-rotational mechanism between 2015-2020.

Therefore, they suggest that Fels landslide is characterized by a bi-planar or multi-planar configuration related to a combination of gravitational and glacial processes.

5.4 Drivers Analysis

Rabus et al. (2022) identified Fels landslide, including lobe b, as a Deep-Seated Gravitational Slope Deformation (DSGSD) and analyzed its sensitiv-

ity to three specific drivers in [30]: surface water from snow-melt and rainfall, locally projected seismic activity of Alaskan earthquakes and lowering of Fels Glacier at the slide toe. They found five displacement patterns that map to independently moving domains of the slide from a surface displacement map-series derived from 1991 to 2016 spaceborne multi-sensor InSAR data. Then, they correlated the corresponding temporal pattern intensities with the suspected drivers using Independent Component Analysis (ICA). The results provide the first evidence that the Fels landslide comprises several independently moving domains that respond differently to the suspected drives as is suggestive of a complex slope deformation.

The landslide domains corresponding to the five independent deformation patterns (IDPs) are believed to be parts of the landslide that are differently affected by the investigated drivers due to lithological, structural and geomorphic characteristics, as well as the depth and orientation of the underlying failure surface. Different landslide domains can have different sensitivities to the primary driver series. In the case of lobe a we're interested in IDP0, IDP1 and IDP3. These areas appear to be outlined by linear features which accord with major trends of lineaments previously discussed. Thus, it can be assumed that the structural geology of the slope controls not only the boundary of the landslide but also domains that have different deformation behaviors relative to a specific primary driver.

The IDP0 is believed to be driven by glacier retreat due to its proximity to the glacier terminus at the toe of the landslide. The relatively consistent rate of the deformation after the 2002 earthquake suggests that this IDP was not affected by the earthquake. However, progressive degradation of rock strength due to deglaciation and debuttressing may have made IDP0 more susceptible to the hydrological driver.

IDP1 is interpreted to be predominantly related to the seismic driver due to high activity of these areas following the Denali earthquake in 2002. This IDP is characterized by larger amplitudes than other IDPs and is spatially

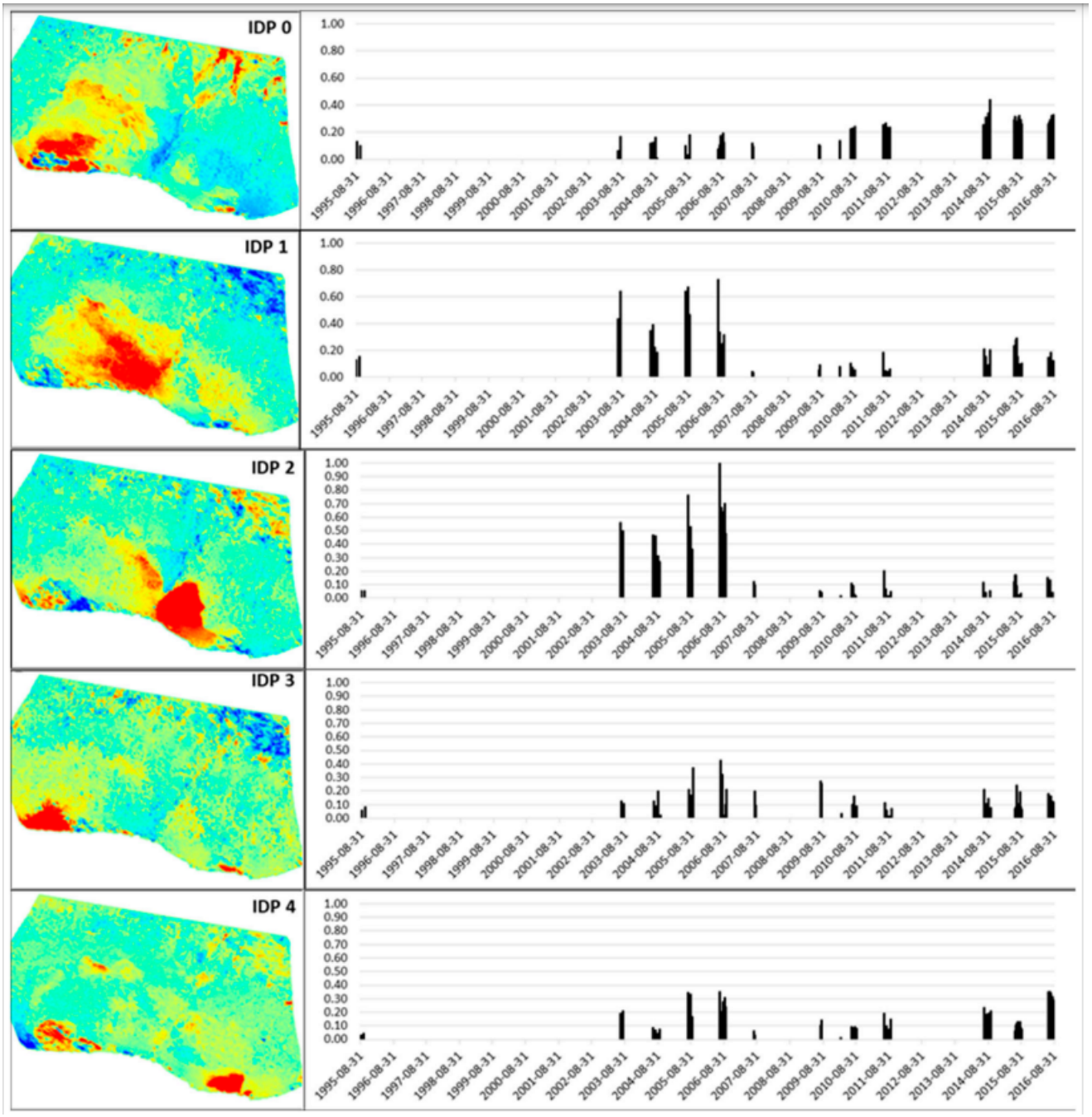


Figure 5.11: Five independent deformation patterns identified by Rabus et al. (2022) with their relative intensities over time. Picture from [30].

more extensive than other signals. Three years after the earthquake, the intensities decreased, indicating that the slope movements had decelerated and an equilibrium reached. The lower intensity movements suggest that the hydrologic cycles are causing some movement, possibly promoted and enhanced by seismic induced damage.

The intensity time series for IDP3 suggest that movements were relatively low before the earthquake and higher in the years following it. This IDP is likely driven by hydromechanical processes that vary on a annual cycle in response snow melt, precipitation and possibly freeze-thaw cycles.

5.5 InSAR Data Comparison

Then the authors compared their observations with the displacement maps derived by Donati et al. (2021) as discussed in the previous sections. They combined RADARSAT-2 data captured in descending geometry with additional data captured in ascending geometry and covering the time period between July 2015 and August 2020. Figure 6.12 show the 3D displacement vectors, the 1D projection of 3D vectors along the descending LoS (Line of Sight) and the vertical displacement component projected to the InSAR LoS. It is evident that, although lobe a is the core active region of Fels slide, lobe a and b in the InSAR projection appear to be of similar strength because the movement direction of less-active lobe b aligns more favorably with the descending geometry LOS. This analysis thus shows that InSAR results for lobe a are not representative as those for lobe b. It's possible that additional, unrecognized IDPs exist in lobe a.

They concluded that Fels slope appears to deform and move as a slope deformation complex. It has a multi-planar, active-passive block configurations and involves a complex three-dimensional failure mechanism with multiple domains affected differently by geological and environmental drivers. The results further highlight the importance of recognizing the complex three-

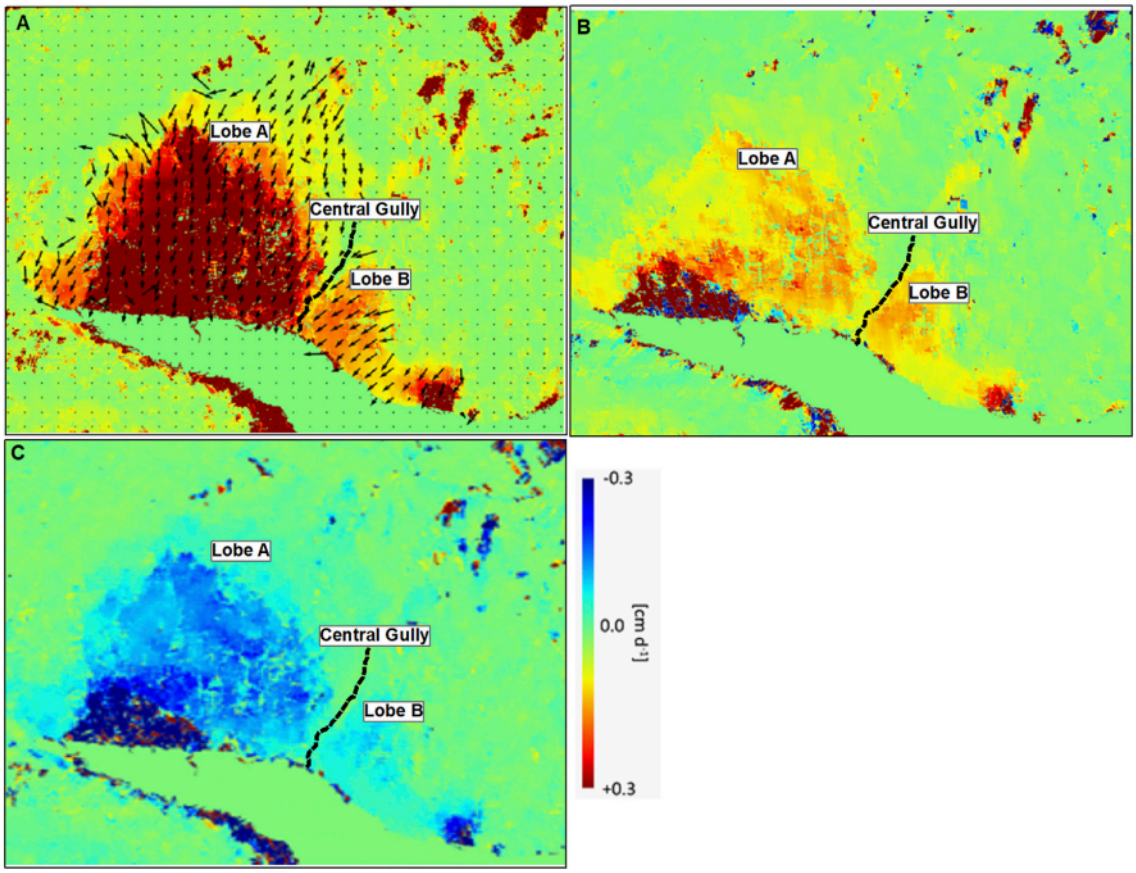


Figure 5.12: Three-dimensional displacements *versus* one-dimensional displacements, only descending orbit geometry of the available data (Section 1.2). **(A)** 3D velocity vectors indicating displacement magnitude (color scale), map direction (arrow aspect), and plunge (arrow length). **(B)** InSAR 1D projection of 3D vectors along the descending LOS. **(C)** Vertical component of 3D velocity vectors. Picture from [31].

dimensional nature of Fels landslide, which may involve different failure mechanisms within different parts of the landslide body, such as translational sliding along foliation, wedge-related sliding and superficial slumping of newly exposed glacial deposits or heavily damaged rock. The lateral growth of the Fels slope failure as the glacier retreats also implies that different zones of movement within the landslide will probably be at different stages of equi-

librium, with mature landslides in the up-valley area and recent or incipient landslides near the current position of the glacier terminus.

Chapter 6

Conclusion

In this work, we experiment the inversion method proposed by Booth et al. (2013) to infer the failure surface geometry of a rock compound landslide in Alaska, the Fels landslide. The available data are the superficial displacement of the landslide acquired by a synthetic aperture radar. From these data we derive the superficial annual average velocity field and we use the classical regularization technique proposed by Tikhonov (1963) to obtain an estimate of the depth of this landslide.

We analyze the results and performance of four different implementations, and we found that the method (Section 4.3) which applies the Balancing Principle and gradient projection algorithm is the most efficient one.

We first experiment the exact method proposed by Booth et al. (2013) which adopts the discrepancy principle and uses CVX solver, a Matlab-based modeling system for convex optimization [27] [28]. Then, we use Balancing Principle instead of the previous principle, since little information about noise affecting data is known. Finally we implement Balancing Principle with gradient projection method proposed by Bonettini et al. (2009) and Newton projection method proposed by Landi and Loli Piccolomini (2008), respectively.

The thickness inferred by all four experiments agree on the general geometric

shape of the sliding surface of Fels landslide. The complex patterns seem to confirm that the Fels landslide is moving as a slope deformation complex as hypothesized by Rabus et al. (2022). The authors also identified three independent deformation patterns (IDPs), which are parts of the landslide that are differently affected by meteorological, seismic, and deglaciation drivers. The IDPs (Figure 5.11) are respectively IDP0, IDP1 and IDP3.

In their first analysis of Fels landslide, Donati et al. (2021) identified three different domains (Figure 5.10): upper domain (UD), lower domain (LD) and central domain (CD).

We consider the thickness inferred by the most efficient method (Figure 4.7) in order to compare the numerical results obtained by the inversion method and the previous knowledge and hypotheses on the structure of Fels landslide.

In the area corresponding to UD, where the "image" seems to be smoother than other parts, a multi-planar block configuration can be identified.
The eastern area corresponding to IDP1 displays more discontinuity features and may be correlated to the fact that this area is predominantly related to the seismic driver.
On the same y-coordinate, in the western part, there is a small region with the same pattern, this may help to identify further IDPs which have not been identified due to measurement constraints (Section 5.5).
Finally the lower part viewed as intersection of LD, IDP0 and IDP3 is the area where least interpretation can be given. The alternation of high values and low values resembles the chaotic pattern of horizontal displacement shown in Figure 3.2.

To summarize, the results of the numerical methods analyzed are consistent with the latest studies on Fels landslide, providing a further evidence that the slope comprises several independently moving domains and the fact that this landslide is better approximated by a "slope deformation complex" model [30].

From a mathematical point of view, further improvements of the inversion method could be achieved by using another regularization function, such as the L1-norm based total variation regularizer proposed by Di Serafino et al. (2020) in [35], or by considering simultaneously different regularization functionals as it is done by Bortolotti et al. (2021) in [36], even if this implies that more efforts are required to solve the optimization (sub)problems.

Since the interpretation of results is not as straightforward as it might be in image processing, it would be valuable to have some synthetic data, which can be considered the ground truth data, from landslides generated by simulation software such as FLAC^{3D} [37] to evaluate quantitatively the performance of the inversion method.

Finally, as real-world data becomes more accessible and abundant with advancement of measuring techniques, one could consider using machine learning techniques, such as artificial neural networks, both to gain insights from data and to improve the efficiency and effectiveness of regularization methods.

List of Figures

Bibliography

[1] Donati D. ; Rabus B. ; Engelbrecht J. ; Stead D. ; Clague J. ; Francioni M. (2021) : A Robust SAR Speckle Tracking Workflow for Measuring and Interpreting the 3D Surface Displacement of Landslides. *Remote Sensing*, 13, 3048.

[2] Handwerger A. L. ; Booth A. M. ; Huang M. H. ; Fielding E. J. (2021) : Inferring the subsurface geometry and strength of slow-moving landslides using 3-D velocity measurements from the NASA/JPL UAVSAR. *Journal of Geophysical Research: Earth Surface*, 126 .

[3] Morton E. Gurtin (1981) : *An Introduction to Continuum Mechanics*, Academic Press.

[4] Ralf Greve : *Dynamics of Ice Sheets and Glaciers*, Lecture Notes 2004/2005, Institute of Low Temperature Science Hokkaido University.

[5] Rasmussen L. A. (1988) : Bed topography and mass-balance distribution of Columbia Glacier, Alaska, U.S.A., determined from sequential aerial photography , *Journal of Glaciology*, Vol. 34, No. 117.

[6] Booth A. M. ; Lamb M. P. ; Avouac J.P. ; Delacourt C. (2013) : Landslide velocity, thickness, and rheology from remote sensing: La Clapière landslide, France. *Geophysical Research Letters*, Vol. 40, 4299-4304.

[7] Zama F. : *Numerical Differential Equations*, Lecture Notes 2021/2022, University of Bologna.

[8] Morigi S. : *Applied Inverse Problems in Imaging: module 1*, Lecture Notes 2021/2022, University of Bologna.

[9] Lanza A. : *Applied Inverse Problems in Imaging: module 2*, Lecture Notes 2021/2022, University of Bologna.

[10] Aster R. C. ; Borchers B. ; Thurber C. H. (2019) : *Parameter Estimation and Inverse Problems*, Elsevier Academic Press.

[11] Ito K. ; Jin B. T. ;Takeuchi T. (2011) : A regularization parameter for nonsmooth Tikhonov regularization, *SIAM Journal on Scientific Computing*, Vol. 33, No. 3, pp. 1415-1438.

[12] Hansen P. C. (2010) : *Discrete Inverse Problems*, SIAM, Philadelphia.

[13] Vogel C. R. (2002) : *Computational Methods for Inverse Problems*, SIAM, Philadelphia.

[14] Porcelli M. : *Matrix and Tensor Techniques for Data Science: module 2*, Lecture Note 2021/2022, University of Bologna.

[15] Nocedal J. ; Wright S. J. (2006) : *Numerical Optimization*, Springer.

[16] Bertsekas D. P. (1999) : *Nonlinear Programming*, Athena Scientific.

[17] Bonettini S. ; Zanella R. ; Zanni L. (2009) : A scaled gradient projection method for constrained image deblurring. *Inverse Problems*, 015002.

[18] Landi G. ; Loli Piccolomini E. (2008) : A projected Newton-CG method for nonnegative astronomical image deblurring. *Numerical Algorithms*, 48, 279-300.

[19] Fuhrmann T. ; Garthwaite M. C. (2019) : Resolving Three-Dimensional Surface Motion with InSAR: Constraints from Multi-Geometry Data Fusion. *Remote Sensing*, 11, 241.

[20] Morlighem M. ; Rignot E. ; Seroussi H. ; Larour E. ; Ben Dhia H. ; Aubry D. (2011) : A mass conservation approach for mapping glacier ice thickness. *Geophysical Research Letters*, Vol. 38, L19503.

[21] Tikhonov A. N. (1963) : Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* 4, pp. 1035-1038.

[22] Larsson S. ; Thomée V. (2003) : *Partial Differential Equations with Numerical Methods*, Springer.

[23] Thomas J. W. (1995) : *Numerical Partial Differential Equations: Finite Difference Methods*, Springer.

[24] Abate M. ; De Fabritiis C. (2015) : *Geometria analitica con elementi di algebra lineare*, McGraw-Hill Education.

[25] Barzilai J. ; Borwein J. (1988) : Two-Point Step Size Gradient Methods. *IMA Journal of Numerical Analysis*, 8, 141-148.

[26] Bertsekas D. P. (1982) : Projected Newton methods for optimization problems with simple constraints. *SIAM J. Control and Optimization*, Vol. 20, No. 2.

[27] Grant M. ; Boyd S. (2008) : Graph implementations for nonsmooth convex programs. *Recent Advances in Learning and Control*, pp 95-110.

[28] Grant M. ; Boyd S. : CVX: Matlab software for disciplined convex programming, version 2.0 beta. *http://cvxr.com/cvx*, September 2013.

[29] Intrieri E. ; Frodella W. ; Raspini F. ; Bardi F. ; Tofani V. (2020) : Using Satellite Interferometry to Infer Landslide Sliding Surface Depth and Geometry. *Remote Sensing*, 12, 1462.

[30] Rabus B. ; Engelbrecht J. ; Clague J. J. ; Donati D. ; Stead D. ; Francioni M. (2022) : Response of a large deep-seated gravitational slope deformation to meteorological, seismic, and deglacial drivers as measured by InSAR. *Frontiers in Earth Science*, 10:918901.

[31] Froude M. J. ; Petley D. N. (2018) : Global fatal landslides occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.*, 18, 2161 - 2181.

[32] Gariano S. L. ; Guzzetti F. (2016) : Landslides in a changing climate. *Earth-Science Reviews*, 162, 227-252.

[33] Dai F. C. ; Lee C. F. ; Ngai Y. Y. (2002) : Landslide risk assessment and management: an overview. *Engineering Geology*, 64, 65-87.

[34] Jaboyedoff M. ; Carrea D. ; Derron M. H. ; Oppikofer T. ; Penna I. M. ; Rudaz B. (2020) : A review of methods used to estimate initial landslide failure surface depths and volumes. *Engineering Geology*, 267, 105478.

[35] Di Serafino D. ; Landi G. ; Viola M. (2020) : ACQUIRE: an inexact iteratively reweighted norm approach for TV-based Poisson image restoration. *Applied Mathematics and Computation*, 364, 124678.

[36] Bortolotti V. ; Landi G. ; Zama F. (2021) : 2DNMR data inversion using locally adapted multi-penalty regularization. *Computational Geosciences*, vol. 25, p. 1215-1228.

[37] Grøneng G. ; Lu M. ; Nilsen B. ; Jenssen A. K. (2010) : Modelling of time-dependent behavior of the basal sliding surface of the Åknes rockslide area in western Norway. *Engineering Geology*, 114, 414 - 422.