Essays on Metric Spaces and Macro-Finance

Thesis submitted by

Hedvig Gál

Doctoral Dissertation

2025

Institute of Economics

Supervisor:

Miklós Pálfia, PhD

Institute of Data Analytics and Information Systems Department of Mathematics
Corvinus University of Budapest

CORVINUS UNIVERSITY OF BUDAPEST
Doctoral School of Economics, Business and Informatics

ESSAYS ON METRIC SPACES AND MACRO-FINANCE

Analysis in Metric Spaces and Macro-Finance aspects

Doctoral Dissertation

HEDVIG GÁL

BUDAPEST, 2025

CORVINUS UNIVERSITY OF BUDAPEST

1. **Title of the Thesis**: *Essays on Metric Spaces and Macro-Finance*

2. **Name, Designation and Institution of the Supervisor:** Miklós Pálfia PhD, Associate Professor, Department of Mathematics, Institute of Data Analytics and Information Systems.

3. **List of Publications:**

   (a) **Journals**

      i. Gál, H., Pálfia, M.(2024).*Convergence of semi-convex functions in CAT(1)-spaces.* Calculus of Variations and Partial Differential Equations. Volume 63, article nr. 213, pp. 1-20.;

      ii. Gál, H., Juhász, A.(2025).*Quantitative Tightening : Theory, Research, and Impact on Selected Emerging Market Economies.* Journal of Central Banking Theory and Practice, Volume 14 (1). pp. 163-181.;

   (b) **Conference Proceedings**

      i. Gál, H., Lovas, A. (2022). *Ergodic Behavior of Returns in a Buy Low and Sell High Type Trading Strategy.* In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds) Mathematical and Statistical Methods for Actuarial Sciences and Finance.MAF 2022. Springer, Cham., pp. 278-283.

      ii. Gál, H. (2022). *A devizaárfolyamok és a részvénypiac volatilitásának előrejelzése a GARCH modellekben*, E-CONOM 11:1, University of Sopron Press, pp. 42-48.

4. **List of Presentations in National/International Conferences and Workshops/PhD Poster session:**

   (a) Mathematical and Statistical Methods for Actuarial Sciences and Finance, University of Salerno (Italy), 20-22 April, 2022.

   (b) International Conference on Empirical Economics, Pennsylvania State University at Altoona (USA), 5 August, 2023.

   (c) 14th Annual Financial Market Liquidity (AFML) Conference, Corvinus University of Budapest (Hungary), 9th-10th November, 2023.

   (d) VOCAL Optimization Conference: Advanced Algorithms, Corvinus University of Budapest (Hungary), 5-7 June, 2024.

   (e) ÚNKP conference, Corvinus University of Budapest (Hungary), 12 June, 2024.

   (f) EUROPT2025 Conference on Advances in Continuous Optimization, University of Southampton, United Kingdom, 29 June - 2 July, 2025

Preface

The framework consists of four chapters and it is written in essay format. The aim is to perform a *theoretical and quantitative analysis* using mathematical concepts and macro-finance approaches.

The connection between each chapter involves making an effort towards the solution of minimization problems: in the specific metric space by the implementation of gradient flow, minimization of financial transaction costs, and monetary intervention by the central bank in order to decrease inflation. Chapter 1. is connected to Chapter 2. and 3. in the context of optimization algorithm setting, which also converges to its minimum. In this context, it means in Chapter 3. the reduction of financial transaction costs using the same Moreau-Yosida approximation on a financial case illustration in Hilbert space such as in Mosco convergence case in CAT(κ)-space. In Chapter 3. the optimization is proved by the Lie-Trotter-Kato formula and the numerical algorithm is performed on a financial dataset. Chapter 4. is connected to Chapter 1., 2., and 3. in a sense of minimization problem to decrease inflation level by the implementation of the Federal Reserve Bank's quantitative tightening program. We can see in every chapter a minimum problem, where the aim is to reach an optimal minimum level. Each chapter refers on the following aspects:

1. **Chapter 1.** - introduces to the reader the main mathematical concepts in order to understand the second chapter of Mosco convergence in **Cartan Alexandrov Topogonov or CAT(1)-space**, where the reader could imagine a geometric shape (very similar to a geometric ball) with it's important trigonometric angles, semi-convex functions, sequence of these functions and the associated gradient flows;

2. **Chapter 2.** - relies on mathematical concepts and characteristics from Chapter 1. and incorporates these mentioned features into the **Mosco convergence of CAT(1)-space**. Particularly, the *Mosco convergence of a sequence of semi-convex lower semi-continuous functions* $f_n : X(-\infty, \infty)$, which converge to $f : X(-\infty, \infty]$. It has been introduced and defined a Mosco convergence of semi-convex lower semi-continuous functions (lsc), which implies a pointwise convergence of Moreau envelopes and it's associated gradient flows in the CAT(1)-space. We used weak compactness to find a weakly convergent subsequence of $J_\tau^{f_n}(x)$ indexed by n. Then, we showed that the limit of this subsequence is actually $J_\tau^f(x)$, and then that the whole sequence strongly converges to $J_\tau^f(x)$. Among others, we also used an Ekeland type of lower estimate on $f_n(x)$, we also proved the same result for Mosco converging sequences of lsc semi-convex L-Lipschitz functions and Mosco converging sequences of lsc semi-convex functions that are uniformly bounded below by 0. The definition of weak convergence and compactness considered here agrees with the usual weak convergence and weak sequential compactness of bounded sets in Hilbert spaces. This minimization approach of a metric distance involves Moreau-Yosida approximation and assiduously focuses on the theory of gradient flows of semi-convex functions established in CAT(κ)-spaces. It refers to the *a priori* estimation of Ohta and Pálfia (2017) in CAT(1)-spaces [90];

3. **Chapter 3.** - considers **Douglas Rachford operator splitting algorithm in Hilbert space**, which converges to a minimum point with the aim of decreasing the financial transaction costs. This dual operator setting is a composition of two monotone functions in the form of bid-ask spread, where exists a binary operation between the two convex functions;

4. **Chapter 4.** - refers to the **Macro-Finance application** in the context of the experience of the Federal Reserve Bank (Fed) related to its policy of Quantitative Tightening (QT) and its spillover effect on BRICS (Brazil, Russia, India, China, South-Africa) and the other selected Emerging Market (EM) economies. It was chosen a sample of countries to examine the impact of the Fed's QT on 10-year government bond yields, between the period of 2012-2022. The result proves that the highest correlation between the long-end yields of the United States and the selected EM has materialized during the first QT (QT1) operation by the Fed between 2017 and 2019 for Peru, Brazil, India and Hungary. It was expected the same behaviour of long-end yields during the second QT (QT2) policy for the selected EM countries.

Acknowledgements

I would like to express my gratitude to those supportive people, who have been part of this scientific journey. First and foremost, thank you to my supervisor, Miklós Pálfia, whose guidance and insightful discussions always added value to my research. I also want to thank the Department of Mathematics for the teaching opportunities and the research group of Analysis in Metric spaces.

I dedicate this doctoral dissertation work to my parents, whose constant immense support throughout my life always motivated and encouraged me.

Also, wish to thank to the Finance and Accounting Department of ELTE University for a supportive community (Mihály Ormos, György Walter, Gabriella Lamanda, Edit Németh, Mária Irén Lakatos, László Kamocsai, Jenő Faró, Gábor Árva) the teaching opportunities, and a new experience in supervising of 20 bachelor students. Thanks to Kata Váradi, who constantly involving me into the research community of the Department of Finance in the Corvinus University of Budapest. I truly appreciate the encouragement of Tibor Illyés, Endre Süli, Gyula Magyarkúti, Gyula Vastag, Tibor Takács, Zsuzsa Réka Huszár, Fülöp András, Helena Naffa, Ágnes Szunomár, Miklós Rosta, Hyungseok Joo, Tamon Asonuma and Robert Faff.

It was a life-changing experience to take part on a "*Statistics in metric space*" workshop in Paris, between 11-13 October, 2023. at ENSAE thanks to the Corvinus University of Budapest and on "*EUROPT2025 Conference on Advances in Continuous Optimization*" at University of Southampton, United Kingdom, between 29 June-2 July, 2025. Thanks to co-author, Attila Juhász (World Bank, Washington) for the collaboration. I am really grateful that could take part in this academic journey of research and teaching, likewise, that could met all of the above listed people.

List of Notation

$dom(\phi)$	Domain of the functional ϕ;
J_τ	Resolvent operator;
$CAT(\kappa)$	Cartan, Alexandrov, and Topogonov metric space[1];
ϕ_τ	Moreau–Yosida approximation of ϕ in $CAT(\kappa)$-space [2];
M^2_κ	Model space of curvature κ and a 2-dimensional Riemannian manifold;
κ	Real number (a parameter), any $\kappa \in \mathbb{R}$;
$Diam_\kappa$	Diameter;
m	Mid-point;
$[x, y]$	Geodesic segment;
$d(x, y)$	Distance of a geodesic segment;
$B_r(x)$	Ball of radius r centered at x in a metric space;
$\bar{\triangle}$	Comparison triangle;
(X, d)	Metric space;
r	Radius of a Ball $B_r(x)$;
env	Moreau-Yosida envelope theorem. See Definition 3.5.1;
n	Number of vertices. See Lemma 1.2.3.;
γ	Closed and rectifiable curve;
$L(\gamma)$	Length of the closed rectifiable curve in a ball (B);
Q	Length space of a sectional curvature not greater than κ;
S	Convex set;
C	Convex subset;
P	Projection point;
$[P_1, P_2]$	Line segment;
S^2	Sphere;
$\mathcal{G}(t, x_0)$	Gradient curve (also: $\xi(t)), S_t(x)$);
P	Diagonal preconditioner;
D, F	Linear operator;
$prox_f(x)$	Proximal operator;
$L(H)$	Bounded linear operator on Hilbert space;
P_f, P_g	Proximal mappings;
R_f, R_g	Reflected proximal mapping, reflectant;

[1]Note: The sectional curvature can be **positive** ($\kappa > 1$), **negative** ($\kappa < 1$) and **flat** ($\kappa = 1$), also the shape of the curvature (or function) can be **convex and semi-convex** ($\kappa > 1$), **concave and semi-concave** ($\kappa < 1$) see *Figure 1.1.*

[2]See Equation 1.27.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction to the mathematical concepts of CAT(1)-space

A specific type of a metric space is called the Cartan-Alexandrov-Toponogov or abbreviated CAT(κ)-space, where the geometric sectional curvature is bounded from above by κ (*real number*). According to the acronym, originally CAT stands for the mathematicians: Élie Cartan, Aleksandr Danilovich Alexandrov, and Victor Andreevich Toponogov. However, Toponogov has never explored curvature bounded from above in his publications, but has made a significant contributions to the theory of conjecture on complete convex surfaces. The acronym itself has been composed by Mikhail Gromov in 1987 [51].

1. **Curvature**: *positive* ($\kappa > 1$), *negative* ($\kappa < 1$) and *flat* ($\kappa = 1$)
2. **Shape of the curve (or function)**: *convex* and *semi-convex* ($\kappa > 1$), *concave* and *semi-concave* ($\kappa < 1$);

Geometric illustration is presented in Figure 1.1.

1.0 History of the Alexandrov geometry

Johann Carl Friedrich Gauss (1777-1855), a German mathematician, contributed in many fields (also spherical geometry) in *the pre-historical era of Alexandrov-geometry.* He recognized that there are geometrical surfaces and there is a lot to do with spherical angles. Later, his disciple, Georg Friedrich Bernhard Riemann (1826-1866) was working on the manifold of non-positive curvature (or the Riemannian manifold). The following mathematicians were also dealing with the mathematical proofs of the Riemannian

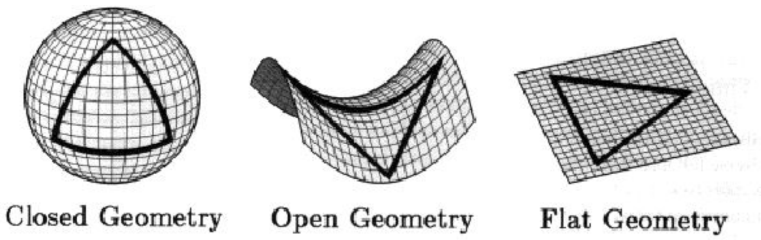

Figure 1.1: Shape and curvature in the CAT(κ)-space. *Source: George Manson University. Note: We are going to deal with positive (κ>1) curvature and semi-convex function.*

manifold: Jacques Salomon Hadamard, Élie Joseph Cartan - his name is also related to the Lie groups and Cartan–Kähler theorem - Herbert Busemann and Abraham Wald. Shortly after the II. World War, Aleksandr Aleksandrov (1912-1999) provided the foundation of Alexandrov geometry and also his students, altogether called *the Russian school* contributed a lot into this field: Yurii Grigorievich Reshetnyak, Yuri Dmitrievich Burago, Igor Nikolaev, Valera N. Berestovskii. Also, Mikhael Gromov popularized the subject and indicated directions of the Alexandrov theory in the modern geometry.

In the last two decades, the theory of convex functions in CAT(1)-spaces has received considerable attention. This theory was first started in the papers [57, 58] of Jürgen Jost (1994, 1998) and [77] Uwe F. Mayer (1998) establishing [31] Crandall-Liggett-type (Crandall and Liggett, 1971) generation theorems for non-linear semi-groups corresponding to semi-convex functions in the CAT(0)-setting. This theorem was also incorporated by Sturm, Ohta, Pálfia, Pazy, Ambrosio, Gigli, Savaré, Reich, Kobayashi, Vázquez and many others. The Crandall-Liggett-type generation theorem provides solutions for differential equations by exponentiating operators. The process of application can be found both in theory and practice. It demonstrates a useful technique of treating functions as points in a vector space. The procedure leads to Euler's implicit method, which describes the behaviour of a resolvent. In this environment, the convergence rate is at least sub-linear, where the outcome is the resolvent formula.

The theory of convex functions has been further expanded [4, 13] in the monographs (Ambrosio L., Gigli, and Savaré G., 2005; Bačák, 2014), where the theory of gradient flows has been studied in great detail. The theory of gradient flows of semi-convex functions established in CAT(1)-space refers to the *a priori* estimation of Shin-ichi Ohta and Miklós Pálfia (2015) in CAT(1)-spaces [89] in their work of *Discrete-time gradient flows and law of large numbers in Alexandrov spaces*. Particularly, in the latter piece of work, the theory of gradient flows considers that the discrete and continuous time flows are leading to a Law of Large Numbers in CAT(1)-spaces.

Without loss of generality, the foundation of Mosco convergence in Hadamard space was established by Kuwae-Shioya (Kuwae and Shioya, 2008) [66] and Bačak (Bačák, 2015) [12]. The Mosco convergence of non-negative convex lower semi-continuous (lsc) functions results pointwise convergence of Moreau envelopes in a CAT(0)-space. These analytical techniques incorporate weak convergence in CAT(1)-spaces and cover asymptotic relations of sequences of such spaces introduced by Kuwae-Shioya (Kuwae and Shioya, 2008) [66], including Gromov-Hausdorff limits.

The measurement of graph similarity using spherical geometry is another application in Alexandrov geometry, which was used by D. Burago, Y. Burago, and S. Ivanov [28] and M. Bačák [13] in CAT(0)-space.

1.1 CAT(κ)-space and Graph comparison geometry

The original concept of CAT(κ) came from Mikhael Gromov, who constructed a specific characterization of non-positive curvature in the metric space. Then, the three authors, Cartan, Alexandrov and Topogonov (CAT) proved that the geodesic triangle satisfies the properties of CAT(κ) inequalities for the cases, when $\kappa > 0$, $\kappa < 0$ and $\kappa = 1$.

Let's take a Model space (M^2_κ), where the sectional curvature is bounded from above by a real positive number: $\kappa > 0$ and this metric space is called a CAT(κ)-space, if we take that $\kappa = 1$, further on we denote it as a CAT(1)-space. The current application involves the same mathematical tools, which are used in the analysis of each metric space in general. Since $\kappa > 0$, we assume that the diameter is exactly $D_\kappa = \frac{\pi}{\sqrt{\kappa}}$ and according to the ***Geodesic framework***, the metric space is D_κ-*geodesic*, where all the pairs of points have a metric distance, which is less than D_κ. Assuming that, the Mosco distance represents a geodesic segment, where the CAT(κ)-space takes a real positive number as mentioned above and in order to describe the spherical environment, we need to recall the definition of a CAT(1)-space from Bridson and Haefliger (1999) from their literature of *Metric Spaces of Non-Positive Curvature* [27]:

Definition 1.1.1 (Definition of CAT(1)-space)**.** Let (X, d) be a CAT(1)-space, where κ is a positive real number. By assumption, the geodesic triangle's $\triangle$ perimeter is smaller than $2D_\kappa$, and a comparison triangle of $\bar{\triangle} \subset M^2_\kappa$. The CAT(1) *inequality* exists if for all $A, B, C \in \triangle$ and for each comparison point $\bar{A}, \bar{B}, \bar{C} \in \bar{\triangle}$, the following inequality holds (Figure 2.1):

$$d(A, B) \leq d(\bar{A}, \bar{B}) \tag{1.1}$$

Then, let's see some additional properties of CAT(κ)-space to understand better this environment according to Bridson and Haefliger (1999) from their book of *Metric Spaces of Non-Positive Curvature* [27]:

- An unique geodesic segment joining each pair of points $x, y \in X$ (Figure 2.1), such as the distance $d(x, y) < D_\kappa$ if $\kappa > 0$) and this geodesic transforms continuously together with its endpoints;
- Every local geodesic in X of length at most $Diam_\kappa$ is *a distance minimizing* geodesic;
- Taking a ball in (X, d) of radius, which is smaller than $Diam_\kappa/2$ and semi-convex, such as any two points in a ball are joined by a unique geodesic segment and this segment is unequivocally contained in the ball;
- Taking a ball $B_r(x)$ in (X, d) with radius (r) less than D_κ is *contractible*;
- Each $\lambda < Diam_\kappa$, where the set is closed and $\epsilon > 0$, there exists $\delta = \delta(\kappa, \lambda, \epsilon)$ and the midpoint (m) of a geodesic segment $[x, y] \subset X$ with $d(x, y) \leq \lambda$ and if $\max[d(x, m'), d(y, m')] \leq 1/2d(x, y) + \delta$, then $d(m, m') < \epsilon$;
- The **Alexandrov-angle** (Figure 2.2) is defined by the monotonicity between the sides of any geodesic triangle in X with distinct vertices, and it is no greater than the angle between the corresponding sides of its comparison triangle $\bar{\triangle} \subset M_\kappa^2$.

1.2 Alexandrov-angle condition of cosine

Definition 1.2.1 (Alexandrov-angle)**.** Let (X, d) be a CAT(1)-space and let γ: $[0, a]$ $\rightarrow X$ and γ': $[0, a'] \rightarrow X$ be the two geodesic segments (γ, γ') emanating from the same point $\gamma(0)$ and $\gamma'(0)$ and the comparison-angle $\angle\ (\gamma(t), \gamma'(t'))$ is a monotone increasing function of $t, t' \geq 0$. The Alexandrov-angle is equal $\angle(\gamma, \gamma')$ to $\lim_{t,t'\to 0} \angle(\gamma(t), \gamma'(t')) = \lim_{t\to 0} \angle(\gamma(t), \gamma'(t))$ according to Proposition 3.1. [27] Bridson and Haefliger (1999).

Therefore, the spherical angle can be written as:

$$\angle(\gamma, \gamma') = \lim_{t\to 0} 2 \arcsin \frac{1}{2t} d(\gamma(t), \gamma'(t)) \tag{1.2}$$

Incorporating the **Law of Cosine** for CAT(1)-space, it leads to the following decomposition [28] (Burago et al., 2001):

$$\cos(c) \leq \cos(a) \cdot \cos(b) + \sin(a) \cdot \sin(b) \cdot \cos(\alpha) \tag{1.3}$$

This polygon contains embedding vectors with the same origin and already defined directions. Let $\overrightarrow{n}$, $\overrightarrow{m}$ and $\overrightarrow{w}$ represent the unit vectors, which go from the center of the Alexandrov-angle to the corners of the CAT(1)-space triangle. Then, $\overrightarrow{n} \cdot \overrightarrow{n} = 1$, $\overrightarrow{m} \cdot \overrightarrow{w} = \cos(c)$, $\overrightarrow{n} \cdot \overrightarrow{m} = \cos(a)$, and $\overrightarrow{n} \cdot \overrightarrow{w} = \cos(b)$. The vectors $\overrightarrow{n} \times \overrightarrow{m}$ and $\overrightarrow{n} \times \overrightarrow{w}$ have the lengths $\sin(a)$ and $\sin(b)$, respectively, and the angle between them is at vertex A (Figure 2.2). Let's implement a *cross product* (vector product), a *dot product* (inner product) and a *Binet-Cauchy identity*, then the proof can be written as follows:

Proof.

$$\begin{aligned}
\sin(a) \cdot \sin(b) \cos(\angle(A)) &= \langle \overrightarrow{n} \times \overrightarrow{m} \rangle \cdot \langle \overrightarrow{n} \times \overrightarrow{w} \rangle \\
&= \langle \overrightarrow{n} \cdot \overrightarrow{n} \rangle \langle \overrightarrow{m} \cdot \overrightarrow{w} \rangle - \langle \overrightarrow{n} \cdot \overrightarrow{m} \rangle \langle \overrightarrow{n} \cdot \overrightarrow{w} \rangle \\
&= \cos(c) - \cos(a) \cdot \cos(b)
\end{aligned} \tag{1.4}$$

□

Apparently, the $\cos(\alpha)$ is a scalar product between the two unit vectors, and both have origin at the vertex of the Alexandrov-angle under the assertion of *isometric transformation* of all geodesics.

We can recall *the list of a priori characterization* by Ambrosio, Gigli, and Savaré (2005) emphasizing the following equations 12.3.1 and 12.3.12-12.3.15 [4], which are useful features to better describe geodesics.

Let (X, d) be a metric space, where x is a constant speed geodesic; and the connection from x to y is a curve. This geodesic of minimal length, whose metric derivative $|x'|(t)$ is constant on $[0, T]$ if $t \in [0, T]$. We can say that X is geodesically complete (or length space) if each couple of points can be connected by a constant speed geodesic [4].

The function d_x is defined in equation 12.3.12 [4], where the fixed $x \in X$ can be denoted as $G(x)$ as a set of all geodesics starting from x and parametrized in some interval $[0, T_x]$, recall that the metric velocity of x is $|x'|= \frac{d(x(t),x)}{t}, t \in (0, T_x)$, therefore:

$$\begin{aligned}
\|x\|_x &:= |x'|, \\
\langle x, y \rangle_x &:= \|x\|_x \|y\|_x cos(\angle(x, y)), \\
d_x^2(x, y) &:= \|x\|_x^2 + \|y\|_x^2 - 2\langle x, y \rangle_x.
\end{aligned} \tag{1.5}$$

If $\lambda > 0$ and $x \in G(x)$, we can say that λx is **a geodesic**:

$$\begin{aligned}
(\lambda x)_t &:= x_{\lambda t}, \\
T_\lambda x &= \lambda^{-1} T_x.
\end{aligned} \tag{1.6}$$

and for each $x, y \in G(x)$ is true:

$$\begin{aligned}
\|\lambda x\|_x &= \lambda \|x\|_x, \\
\langle \lambda x, y \rangle_x &= \langle x, \lambda y \rangle_x = \lambda \langle x, y \rangle_x
\end{aligned} \tag{1.7}$$

where, the restriction of a geodesic is still a geodesic and $x \sim y$ if $\epsilon > 0$ such that

$x|_{[0,\epsilon]} = y|_{[0,\epsilon]}$. Here, **the $x \sim y$ is an equivalence relation if two geodesics are isometric as well**. A composition of isometries and the inverse of an isometry is again an isometry [69], therefore being isometric is an equivalence relation (Lee J.M., 2018). The quotient space of the unit square, for example $[0,1] \times [0,1]$ by the equivalence relation [70] generated by the following relations (Lee J.M., 2011):

$$\begin{gathered}(x,0) \sim (x,1) \;\; for \;\; 0 \leq x \leq 1 \\ (0,y) \sim (1,1-y) \;\; for \;\; 0 \leq y \leq 1\end{gathered} \tag{1.8}$$

Theorem 1.2.1 (An abstract notion of Tangent cone)**.** If $x, y : [0,T] \to X$ are two geodesics starting from x:

$$d_x(x,y) = \lim_{t \downarrow 0} \frac{d(x_t, y_t)}{t} = \sup_{t \in (0,T]} \frac{d(x_t, y_t)}{t} \tag{1.9}$$

The function d_x defined by 1.5 is a distance on the quotient space and the completion on the quotient space is called the tangent cone $Tan_x\ X$ at the point x.

Definition 1.2.2 (Tangent-cone)**.** The Tangent cone $(G_x/{\sim})$ is specified as a Gromov-Hausdorf limit of the pointed sequence in CAT(1)-spaces and its proof incorporates the upper-angle condition of cosine in 12.3.15 (Ambrosio, Gigli and Savaré, 2005).

Let $G(x)$ be the set of geodesics, which has a constant speed, and starting from x, where each is parameterized by taking respective values from an interval $[0,T_x]$. Then, the metric velocity can be defined as an inner product incorporating ***the Alexandrov-angle condition of cosine*** (Ambrosio, Gigli, and Savaré, 2005):

$$\langle (s,\gamma),(t,\eta) \rangle := st \cdot \cos(\angle_x(\gamma,\eta)) \tag{1.10}$$

Let the function d_x be the distance in the space, where $G_x/{\sim}$ is a Tangent-cone at the point x [4] (Ambrosio, Gigli, and Savaré, 2005):

$$\begin{gathered}d_x^2((s,\gamma),(t,\eta)) := s^2 + t^2 - 2st\cos\angle_x(\gamma,\eta) \\ \lambda \cdot (s,\gamma) := (\lambda s, \gamma) \\ \lambda \cdot (t,\eta) := (\lambda t, \eta)\end{gathered} \tag{1.11}$$

Definition 1.2.3 (Spherical Law of Cosine in CAT(1)-spaces)**.** Let (X,d) be the CAT(1)-space of a two-dimensional vector space of S^2, where *the distance is defined by an angle cos(α)* from which the two unit vectors (a scalar product) have a dispose of the same origin and satisfy *the condition of triangle inequality*. Without loss of generality, the Law of Cosine with **identical geodesic lengths (a, b and c) and with vertices (A, B and C)** can be written in the following way:

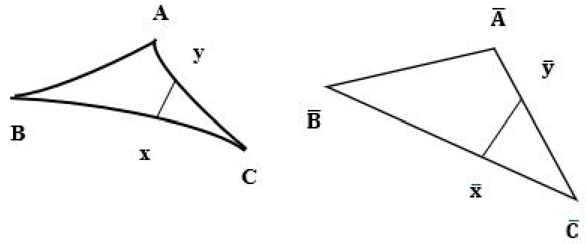

Figure 1.2: CAT(1) inequality of a geodesic (x, y) and a CAT(0) comparison triangle

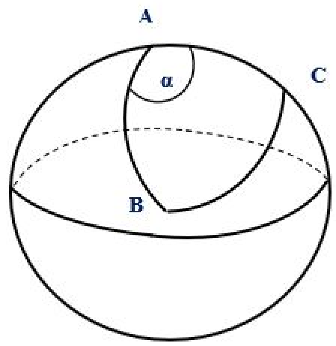

Figure 1.3: Geodesic triangle with Alexandrov-angle in the CAT(1) metric space

Definition 1.2.4 (Cross-product)**.** The cross-product of two real vectors, which is in notation $a \times b$ a binary operation of two vectors is defined only in three-dimensions (Wilson, 1901) [22]. The cross product is characterized as a vector c that is *perpendicular* (orthogonal) to both a and b, with a direction given by the right-hand rule and a magnitude equal to the area of the parallelogram that the vectors span (Zill and Cullen, 2006) [108].

$$c^2 = a^2 + b^2 - 2ab \cdot cos(\alpha) \tag{1.12}$$

Proof.

$$\left\|\overrightarrow{a - b}\right\|^2 = \left\|\overrightarrow{a + b}\right\|^2 - 2 \cdot \left\|\overrightarrow{a \times b}\right\| \cdot cos(\alpha) \tag{1.13}$$

□

where,
- α is the angle between a and b in the plane (between 0 and 180 degree);
- $\|a\|$ and $\|b\|$ are the magnitudes of the given vectors a and b;
- n $=$ 2 is a unit vector perpendicular to the plane including a and b, with direction such that the ordered set (a, b, n) is positively oriented.

The angle between geodesics refers to the constant speed of geodesics in Remark 12.3.5. (Ambrosio, Gigli and Savaré, 2005), when $\|x\|:= |x'|$. Then, the metric derivative could estimate the norm of the velocity field, and the triangle inequality by involving angles possess the following spherical aspects [4]:

$$\angle(A_B^C)_{S^2} \leq \angle(B_A^C)_{S^2} + \angle(C_A^B)_{S^2}. \tag{1.14}$$

Definition 1.2.5 (Spherical angle)**.** According to the definition of Bridson and Haefliger (1999) the spherical angle between two minimal great arcs issuing from a point of S^n (n-dimensional sphere), with initial vectors u and v say, is the unique number $\alpha \in [0, \pi]$ such that $cos(\alpha) = (\mathrm{u} \mid \mathrm{v})$ [27].

Definition 1.2.6 (Spherical triangle)**.** A spherical triangle in S^n (n-dimensional sphere) consists of a choice of three distinct points (its vertices) A, B and $C \in S^n$, and three minimal great arcs (its sides) one joining each pair of vertices [27].

Definition 1.2.7 (Vertex angle)**.** The vertex angle at C is defined to be the spherical angle between the sides of $\triangle$ joining C to A and C to B [27].

Proposition 1.2.1 (2.9 [27])**.** The spherical angle between two geodesic segments [C, A] and [C, B] in S^n is equal to the Alexandrov angle between them.

By decreasing one angle, the geodesic distance will also decrease and vice versa. The perimeter is bounded by $2D_\kappa$, which ensures that the geodesic triangle is included with its geodesic points (x, y). This geodesic polygon possesses a length minimizing curve including a sequence of points on that geodesic segment and the semi-convexity holds among the initial and final point of a curve.

In the context of polygons, we can also speak about the quadruple (or 4-point) graph comparison geometry (Figure 1.4):

Definition 1.2.8 (Quadruple comparison)**.** Let (X, d) be a CAT(1)-space, where the *geodesic quadrangle's* $\tilde{\varnothing}$ perimeter is smaller than $2D_\kappa$. By taking a comparison quadrangle of $\varnothing$ with its comparison points of $A, B, C, D \in \varnothing$ for the Euclidean space in association with the geodesic quadrangle's points of $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D} \in \tilde{\varnothing}$, then the following CAT(1) *inequality* holds:

$$d(A, D) + d(D, C) \leq d(\tilde{A}, \tilde{C}) \tag{1.15}$$

Based on Alexandrov's lemma for the quadrangle $\varnothing(A, B, C, D)$, where A and C are separated by a geodesic line $d(B, D)$, we obtain [28]:

$$d(B, D) \leq d(\tilde{B}, \tilde{D}) \tag{1.16}$$

$$d(A, B) \leq d(\tilde{A}, \tilde{B}) \tag{1.17}$$

$$d(B, C) \leq d(\tilde{B}, \tilde{C}) \tag{1.18}$$

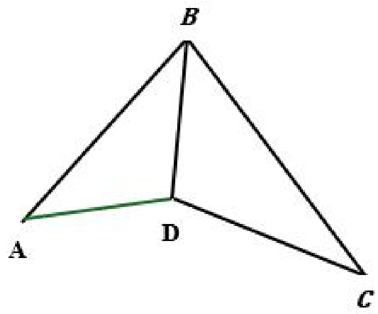

Figure 1.4: Comparison quadrangle in the Euclidean space

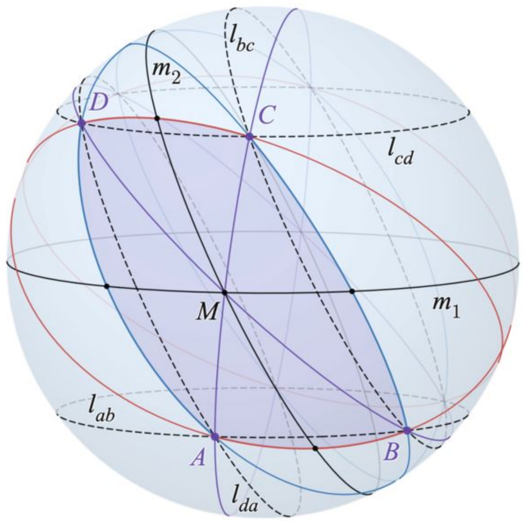

Figure 1.5: Geodesic quadrangle in the CAT(1)-space. *Source: Desmos/geometry https://www.desmos.com/geometry/e8rsvot3to.*

The opposite sides and opposite angles are *congruent*, and as the angles increase, similarly the distance also increases in the CAT(1)-space, then for angles: $\alpha < \tilde{\alpha}$; $\beta < \tilde{\beta}$; $\gamma < \tilde{\gamma}$; and $\delta < \tilde{\delta}$. Alexandrov's Lemma in the nutshell describes all the properties, which a quadruple geodesic contains in the theory of curvature bounded from above.

Figure 1.4. illustrates the Euclidean model space, where according to Alexandrov's lemma 1.5.1., the following facts are emphasized (Alexander, Kapovitch and Petrunin, 2019) [1]:

1. Let A, C, B and D be points in the metric space such that $D \in [A, C]$;

2. $\angle(B^A_C) \leq \angle(B^A_D) + \angle(B^D_C)$;

3. The same angle condition holds for the spherical models with the addition of $d(B, D) + d(B, C) + d(A, C) < 2\pi$.

Similarly, Figure 1.5 in the CAT(1)-space depicts the geodesic quadrangle (or parallelogram) of $\tilde{\varnothing}$ $(\tilde{A}, \tilde{C}, \tilde{B},\ \tilde{D})$. The Law of parallelogram indicates geodesic triangels in CAT(1)-space, where geodesics are thinner than in Euclidean space. These properties indicate distances and angel conditions of inequality, which serves as a comparison analysis with the Euclidean space or other selected spherical model space. In the Euclidean space, the geometric features are congruent, if they have the same size and shape (*such as the congruent triangles*)[1].

Example: Let us choose a quadrilateral parallelogram with equal opposite sides (Figure 1.4):

$$(1 + \cos\ d_1) \cdot (1 + \cos\ d_2) = (\cos\ a + \cos\ b)^2 \tag{1.19}$$

where d_1 and d_2 are the diagonal lengths. The sidelenghts are equal with:

$$a = b \cos C + d \cos B \tag{1.20}$$
$$b = d \cos A + a \cos C \tag{1.21}$$
$$d = a \cos B + b \cos A \tag{1.22}$$
$$\tag{1.23}$$

Assuming $d_1 = d_2 = d$, we obtain:

$$(1 + \cos\ d) = (\cos\ a + \cos\ b) \tag{1.24}$$

And,

$$\det(ABC)^2 \cdot \det(ACD)^2 = \det[\det(ABC) \cdot \det(ACD)]^2 \tag{1.25}$$

The quadrilateral parallelogram contains two triangles: $\varoslash(ABCD) = \triangle(ABC) + \triangle(ACD)$ with equal sidelengths. Berg and Nikolaev (2015) elaborated the aspects of quadruple comparison including Reshetnyak's Majorization Theorem for the CAT(1)-space [18]. In the following, let's have a look at the Reshetnyak's Majorization Theorem:

Theorem 1.2.2 (Reshetnyak's Majorization Theorem)**.** Let B be a ball, $B \subset X$ in a metric space of curvature bounded from above, and γ be a closed rectifiable curve in B. If $\kappa > 0$, then the length of the closed rectifiable curve $L(\gamma) \leq \frac{2\pi}{\sqrt{\kappa}}$. Then, there exists a closed convex set D in the κ plane bounded by a closed curve Γ and a non-expanding map f of D into B such that the restriction $\tilde{f}$ of f on Γ is a length-preserving map onto γ in M^2_κ [28]. Since, for any closed curve Γ in a CAT(κ) space, there exists a convex region D in M^2_κ, and an associated map f such that D majorizes Γ under f.

See later in Theorem 2.3.1 Reshetnyak's Majorization Theorem in the context of Mosco convergence.

If the statement of the theorem holds for every (sufficiently short) closed curve γ in a length space Q, then Q is a space of curvature not greater than κ (Burago et al., 2001; see in Section 9.1.23). This theorem works for Hadamard space and CAT(1)-setting,

[1]Oxford Dictionary

because there exists a convex region D in M_κ^2, and an associated map f such that D majorizes Γ under f.[2]

Similarly, in Lemma 1.13 (Section 1.5), Toyoda Tetsu from Kogakuin University in Tokyo (2023) used Reshetnyak's Majorization Theorem [101] in the context of quadrilateral parallelogram in CAT(κ)-space below in the equation of 1.27. Prior that, let's introduce a model space.

The real number κ of M_κ^2 is complete, two-dimensional Riemannian manifold of constant Gaussian curvature κ, and by d_κ the distance function on M_κ^2. Let D_κ be the diameter of M_κ^2 [101]:

$$Diam_\kappa = \begin{cases} \frac{\pi}{\sqrt{\kappa}} & \text{if } \kappa > 0; \\ \infty & \kappa \leq 0. \end{cases} \tag{1.26}$$

Lemma 1.2.3 (Graph comparison of a quadrilateral parallelogram in Section 1.5, Lemma 1.13 in [101])**.** Suppose a quadrilateral parallelogram in the Euclidean space of $A, B, C, D \in X$ and a comparison quadrilateral parallelogram in CAT(κ)-space $A', B', C', D' \in M_\kappa^2$. Let $\kappa \in \mathbb{R}$ and let (X, d_x) be a $Cycle_n(\kappa)$-space, which is an isometric embedding into CAT(κ)-space and equivalent to being CAT(κ). Here, n is a number of vertices ($n = 4$) and we use the same notation of vertices A, B, C and D as in the previous 1.25 equation. The sum of distances between each pair of vertices in CAT(κ)-space must be smaller than the $2Diam_\kappa$ diameter bound:

$$\begin{gathered} d_\kappa(A', B') + d_\kappa(B', C') + d_\kappa(C', D') + d_\kappa(D', A') < 2Diam_\kappa; \\ d_x(A, B) \leq d_\kappa(A', B'); \\ d_x(B, C) \leq d_\kappa(B', C'); \\ d_x(C, D) \leq d_\kappa(C', D'); \\ d_x(D, A) \leq d_\kappa(D', A'); \\ d_x(A, C) \leq d_\kappa(A', C'); \end{gathered}$$

In this context, every distance between two vertices d_κ (with its endpoints) in CAT(κ)-space is considered as a unique geodesic segment.

[2]Further reference see in Theorem 2.7.1.(Burago et al., 2001).

1.3 Theory of Convex Analysis

The convex optimization theory in the CAT(0)-setting is described in-depth in *Convex Analysis and Optimization in Hadamard Spaces* [13] by M. Bačák (Bačák, 2014). In the first place, let's see some concepts in the convex analysis and design: geodesic convexity, projection onto convex subset and the Evolution Variational Inequality.

1.3.0 Concepts of Optimum Design

Definition 1.3.1 (Convex set)**.** A convex set S is a collection of points having the proper shape if P_1 and P_2 are any points within S, then the line segment of $[P_1,P_2]$ is also in C. This is a necessary and sufficient condition for convexity of the set S. (Arora, 2017) [9]. See Figure 1.6.

Definition 1.3.2 (Geodesic convexity)**.** A functional f is called *convex* on a curve $\gamma : t \in [0,1] \to \gamma_t$, if the curvature shape between the initial and final point of the curve is convex or semi-convex γ [4] (Ambrosio, Gigli, Savaré, 2005):

$$f(\gamma_t) \leq (1-t)f(\gamma_0) + tf(\gamma_1) \tag{1.27}$$

Additional fact is that, the functional $f : X \to \mathbb{R}$ is convex or semi-convex for the geodesic path $\gamma : I \to X$ and it is parameterized proportionally to the arc length [4].

Definition 1.3.3 (Projection onto Convex subset)**.** Let (X,d) be the CAT(1)-space and let $C \subseteq X$ be a complete C-convex subset within the (X,d) metric space, where the property of an obtuse-angle (an angle between 90 and 180 degree) in Figure 1.7 [98]) holds if the distance of an initial point is $\frac{\pi}{2}$ metric inter-space (projection to a nearest point) within the C-convex subset [3]. The orthogonal projection or simply projection on C-convex subset of CAT(1)-space is given by a map $\pi : (X,d) \to C$ [27]. See Figure 1.7.

1.3.1 Examples of Convex Analysis

By the replacement of X [27] by $V = \{x \in X | d_C(x) < \frac{D_\kappa}{2}\}$ in the Exercise 2.6., we can obtain the following properties (Bridson and Haefliger, 1999) for the Proposition 2.4.:

1. $\forall x \in V$ there exists a unique point $\pi(x) \in V$, such that $d(x,\pi(x)) = d(x,C) := \inf_{y \in C} d(x,y)$;

[3] The two common methods used to solve oblique triangles are the law of sines and the law of cosines in trigonometry.

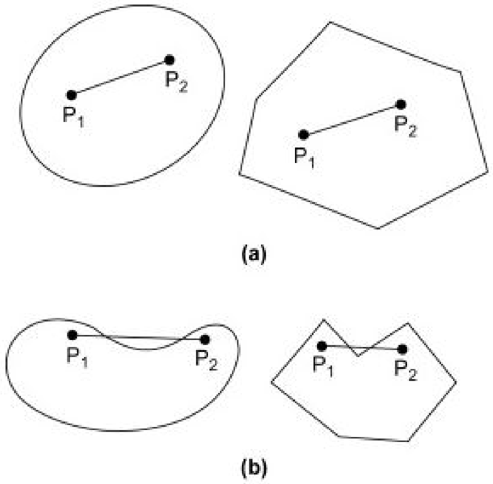

Figure 1.6: Examples of (a) convex and (b) non-convex sets. *Source: Arora, J., Introduction to Optimum Design, Elsevier, 2017.*

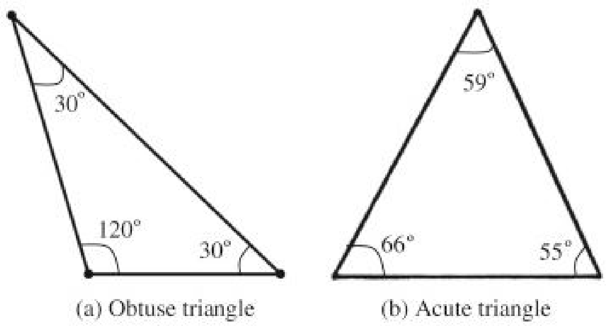

Figure 1.7: Obtuse and acute triangle. *Note: Obtuse angle is any angle greater than 90 degree and Acute angle is any angle less than 90 degree. Source: Sarhan M. Musa, Fundamentals of Technical Mathematics, Elsevier, 2015.*

2. If x', it belongs to the geodesic segment $[x, \pi(x)]$, then $\pi(x') = \pi(x)$;

3. The property of obtuse-angle holds as a $y \in C$, and $x \notin C$, than the $\angle_{\pi(x)} \geq \frac{\pi}{2}$ as a $y \neq \pi(x)$;

4. Continuous homotopy can be obtained from the identity map of X to π, as a retraction of X onto C, the metric distance does not increase and starts from a point projecting homotopy-equivalence onto a geodesic segment. In Figure 1.8. the given P projected to a point P' on a convex subset will get closer to any other point Q (Bagnell, 2014) [14].

Another important concept in the theory of convex and semi-convex analysis is convergence, such as the weak convergence. The weak convergence or weak topology can be identified by the asymptotic center in Lemma 28. [62] (Kell, 2014). Let's see the definition of the weak convergence:

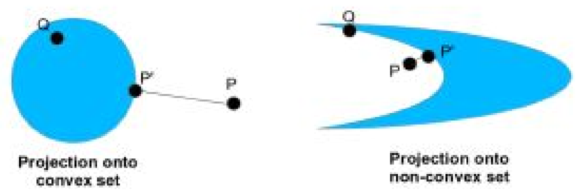

Figure 1.8: Projection onto Convex and Non-Convex subset. *Source: Drew Bagnell, Convex Optimization and Subgradients, Carnegie Mellon University, 2014.*

Definition 1.3.4 (Weak convergence)**.** A net $x_i \in X_i$ *weakly converges* to a point $x \in X$, that is $x_i \xrightarrow{w} x$, if for any net of π-convex geodesic segments $\gamma_i \subseteq X_i$ strongly converging to a π-convex geodesic segment $\gamma \in X$ with $\gamma(0) = x$, such that, $d_i(\gamma_i, x_i) := \inf_{z\in\gamma_i} d_i(z, x_i) < \pi/2$, we have that $P_{\gamma_i}(x_i)$ strongly converges to x. Strong convergence of γ_i to γ means that $\gamma_i(t) \to \gamma(t)$ for each $t \in D(\gamma)$.

The following Proposition 3.8. refers to the literature of *Uniformly convex metric spaces* by Kell (2014):

Lemma 1.3.1 (Proposition 3.8.(Kell, 2014))**.** Let $\mathrm{diam}(X) < \pi$. Let $f : X \to (-\infty, \infty]$ be a quasi-convex lsc function (Kell, 2014). Then, f is weakly lsc. In particular, $x \to d^2(a, x)$ is weakly lsc on $B_a(\pi/2)$ [62].

Definition 1.3.5 (Mosco convergence)**.** A sequence of lsc functions $f_n : X \mapsto \overline{\mathbb{R}}$ said to converge to $f : X \mapsto \overline{\mathbb{R}}$ in the sense of Mosco if, for any $x \in X$, we have

(M1) $f(x) \leq \liminf_{n\to\infty} f_n(x_n)$ whenever $x_n \xrightarrow{w} x$,

(M2) there exists an $(y_n) \subseteq X$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$.

Proposition 1.3.2 (Ekeland principle for a Mosco convergent sequence of lsc semi-convex functions, a bounded case)**.** Let X be a CAT(1)-space with $\mathrm{diam}(X) < \pi$. Given $x_0 \in X$ and a sequence of lsc λ-convex functions $f_n : X \to (-\infty, \infty]$ that is Mosco converging to $f : X \to (-\infty, \infty]$, there exists $\alpha, \beta \geq 0$ such that:

$$f_n(x) \geq -\alpha d(x, x_0) - \beta$$

for all $x \in X$ and $n \in \mathbb{N}$.

The function $f(\cdot, (x_i))$ has a *unique minimizer*, and this minimizer is called *the asymptotic center*. The definition of weak sequential convergence (Kell, 2014) clarifies the importance of the asymptotic center [62]. The sequence $(x_i)_{i\in\mathbb{N}}$ weakly converges to a point x, if x is *the asymptotic center* for each and every subsequence of x_n, in notation: $x_i \xrightarrow{w} x$. The sublevel sets of a function are closed, bounded and convex, particularly weakly-closed, according to the evidence in the proof of Lemma 2.8 (Kell, 2014).

In the current chapter, Definition 1.3.4. elaborates on the weak convergence setting in the CAT(1)-space under the condition of π-convexity and Lemma 1.3.1 regarding the $\operatorname{diam}(X) < \pi$. The previously mentioned Definition 1.3.4. clarifies the basic properties of Mosco convergence in CAT(1)-space, and Proposition 1.3.2. introduces the Ekeland principle for a Mosco convergent sequence of lsc semi-convex functions in the bounded case.

Firstly, Ohta and Pálfia have developed a theory of gradient flows in *discrete* [89] (Ohta and Pálfia, 2015) and *continuous-time* [90] (Ohta and Pálfia, 2017) for the lsc semi-convex functions on CAT(1)-spaces. If $\kappa > 0$, the metric $d(\cdot,\cdot)$ of a CAT(κ)-space can be rescaled to $\frac{1}{\sqrt{k}}d(\cdot,\cdot)$, so that it becomes a CAT(1)-space with this new metric. Using the *Moreau-Yosida* resolvent $J_\tau^\phi(x)$ for the lsc λ-convex function $\phi : X \mapsto (-\infty,\infty]$ and $\tau > 0$, one can construct the gradient flow ($S(t)x_0 = \mathcal{G}(t,x_0) := \xi(t)$) as

$$S(t)x_0 = \lim_{n\to\infty} \left(J_{t/n}^\phi\right)^n (x_0) \tag{1.28}$$

for the initial point $S(0)x_0 := x_0$ and $t > 0$. This curve $\xi(t) = S(t)x_0$ is characterized as the unique solution [83] to the *Evolution Variational Inequality* (Muratori and Savaré, 2020):

$$\frac{1}{2}\frac{d}{dt}[d^2(\xi(t),y)] + \frac{\lambda}{2}d^2(\xi(t),y) + \phi(\xi(t)) \leq \phi(y)$$

for all $y \in dom(\phi)$ and almost all $t > 0$.

Secondly, the variational convergence for sequences of functions in CAT(0)-spaces has been considered, see Jost [57] (Jost, 1994) and then Kuwae and Shioya [66] (Kuwae and Shioya, 2008), in which was introduced the concept of *Mosco convergence* for sequences of convex functions in a CAT(0)-space extending the pioneering works of Mosco [81] (Mosco, 1994) and Dal Maso [34] (Dal Maso, 1993). See also Attouch [10] (Attouch, 1984) for a very detailed approach to the Banach and Hilbert settings. In particular, it has been established in the work of Kuwae [66] (Kuwae and Shioya, 2008), that Mosco convergence of lsc convex functions implies the convergence of the corresponding Moreau-Yosida resolvents and in turn convergence of their gradient flows provided that all functions are bounded from below. This latter requirement has been relaxed in the work of Bačák [12] (Bačák, 2015) and established the convergence through the gradient flow (1.28), also in the case of limits of sequences of metric spaces. They considered the general notion of an *asymptotic relation* of metric spaces, which includes, among others, Gromov-Hausdorff limits. An essential feature of these approaches is the concept of *weak convergence* in the CAT(0)-space, where the closed and bounded sets turn out to be weakly sequentially compact (Bačák, 2015). The concept of weak convergence and compactness carries over to the CAT(1)-setting as it has been investigated by a number of authors along with the study of (semi-)convex functions [62, 106] (Kell, 2014; Yokota, 2016).

In Chapter 2. we will generalize the convergence of resolvents and gradient flows of Mosco convergent sequences of lsc semi-convex functions to the CAT(1)-setting from the CAT(0) by taking into account the convex functions case [66, 12, 13] (Kuwae and Shioya, 2008; Bačák, 2014, 2015). In order to utilize weak convergence, thus to make sense of Mosco convergence, we need to assume some diameter bounds on our spaces in the most general cases of arbitrary lsc semi-convex functions. Under somewhat stronger assumptions of uniform lower boundedness or Lipschitz continuity of our function sequence, we are able to prove convergence without diameter restrictions. Also, we extend to the case of asymptotic relations, by generalizing the required weak convergence and compactness results to the CAT(1)-setting, which may be of independent interest.

In [12] (Bačák, 2015) the author utilizes local and global slope estimates for lsc functions in the CAT(0)-setting established in [4] (Ambrosio, Gigli, Savaré, 2005). Pálfia and Ohta in [90] (Ohta and Pálfia, 2015) established more direct estimates of error terms by minimizing movements related to gradient flows for semi-convex functions compared to those in [4] (Ambrosio, Gigli, Savaré, 2005). We utilize these estimates in order to control error terms of minimizing movements converging to gradient curves. This approach is more direct than the one available for CAT(0)-spaces in [12] (Bačák, 2015), thus providing a simpler approach. Also, we rely on Lipschitz estimates of Kuwae (2015) [64] established for Moreau-Yosida resolvents locally in CAT(1)-spaces. The local nature of our analysis is one of the essential differences between the Hadamard space and CAT(1)-setting.

Before moving onto the next chapter, we will fix some notations. Given a real number $\kappa > 0$, then M_κ^2 is obtained from the sphere S^2 by rescaling the distance function by the constant $1/\sqrt{k}$. For the definition of CAT(κ)-spaces we refer to the textbook [27] (Bridson and Haefliger, 1999). A subset A of a CAT(κ)-space (X, d) is said to be λ-convex for $\lambda > 0$ if for any two points $x, y \in A$ with $d(x, y) < \lambda$ can be joined by a unique minimal geodesic denoted by $[x, y]$ contained in A, where all CAT(κ)-spaces are assumed to be complete.

Definition 1.3.6 (semi-continuity)**.** A function $\phi : X \mapsto \overline{\mathbb{R}}$ with its effective domain $D_\phi := \{y \in X | \phi(y) < \infty\}$ is *lower semi-continuous* (lsc) at $x_0 \in D_\phi \subseteq X$, if

$$\liminf_{x \to x_0} \phi(x) \geq \phi(x_0).$$

The lsc function $(\phi(x))$ is continuous from below and when the $\phi(x)$ is greater than $\phi(x_0)$ point (see Figure 1.9). Similarly, one can define weak lsc at $x_0 \in D_\phi \subseteq X$ if

$$\liminf_{x \xrightarrow{w} x_0} \phi(x) \geq \phi(x_0).$$

A function $\phi : X \mapsto \overline{\mathbb{R}}$ is called lower semi-continuous (lsc) on D_ϕ, if ϕ is lsc at any

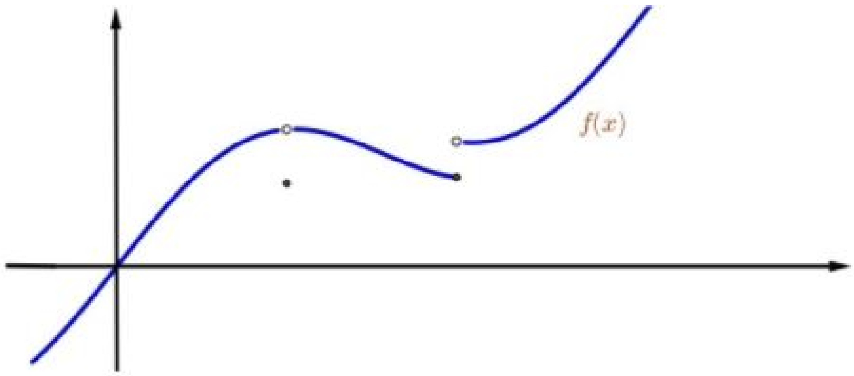

Figure 1.9: Lower semi-continuous function ($f(x) = \phi(x)$). *Source: Lafferriere, and Nguyen, Portland State University.*

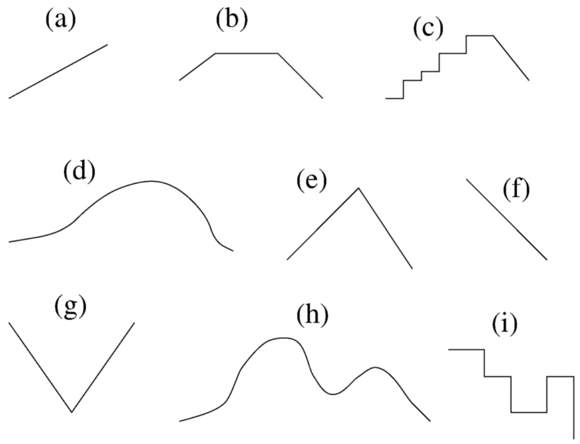

Figure 1.10: Examples of semi-convex functions (a)-(f) and non-semi-convex functions (g)-(i). *Source: Paul Morris, Solving and learning a tractable class of soft temporal constraints: Theoretical and experimental results, DBLP, 2007.*

point in D_ϕ. It is well known that a function $\phi : X \mapsto \overline{\mathbb{R}}$ is lsc iff all sub-level sets are closed, iff its epigraph is closed.

Definition 1.3.7 (semi-convexity)**.** A function $\phi : X \mapsto \overline{\mathbb{R}}$ is *semi-convex* see Figure 1.10, more precisely λ-*convex* for a $\lambda \in \mathbb{R}$, if

$$\phi(\gamma(t)) \leq (1-t)\phi(\gamma(0)) + t\phi(\gamma(1)) - \frac{\lambda}{2}(1-t)td^2(\gamma(0), \gamma(1)) \tag{1.29}$$

along any geodesics $\gamma : [0, 1] \mapsto X$.

More generally, ϕ is *quasi-convex* if its sublevel sets are convex.

1.4 Theory of Gradient flow

Ohta-Pálfia (2017) constructed through the discrete variational application the gradient flow of a lsc and semi-convex function in the CAT(1)-space by the implementation of Moreau-Yosida approximation, see 1.1 equation as a reference (Ohta-Pálfia, 2017), where point x_τ is reaching infimum via the approximation $\xi(\tau)$, such as the gradient curve $\xi(0) = x$.

The gradient curve of ϕ is semi-convex if it is λ-convex for $\lambda \in \mathbb{R}$. In a nutshell, ϕ is lsc and a semi-convex function on the metric space (X, d).

Ambrosio, Gigli and Savaré (2005) in their book of *Gradient Flows in Metric Spaces and in the Space of Probability Measures* have gradual steps about the *Proofs of the Convergence Theorems* [4]. The main convergence itself in their estimations consists of four steps in Chapter 3:

1. A single minimization problem of the scheme, where the resolvent operator provides all the solutions - see 2.0.4 in Ambrosio, Gigli and Savaré (2005);

2. Stability estimates are derived for discrete solutions;

3. It leads to Proposition 2.2.3 by the compactness;

4. Convergence is obtained by combining the *a priori* energy estimates with the gradient properties of the relaxed slope.

The theory of gradient flows in unique metric space has implementation in heat flow as gradient flow of the relative entropy in the (L^2-) Wasserstein space, initiated by Jordan, Kinderlehrer and Otto [55], is known as a useful technique in partial differential equations (see [92], [104], [4], [6] among many others), and has played a crucial role in the recent remarkable development of geometric analysis on metric measure spaces satisfying the (*Riemannian*) *curvature-dimension condition* (see [103], [45], [47], [3], [5], [39]).

Corollary 1.4.1. Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume that there is a sequence $\{t_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} t_n = \infty$ and $\{\xi(t_n)\}_{n\in\mathbb{N}}$ converges to a point $\bar{x}$. Then $\bar{x}$ is a stationary point of ϕ and $\lim_{t\to\infty} \phi(\xi(t)) = \phi(\bar{x})$.

Ohta and Pálfia (2017) followed a standard strategy of constructing a gradient flow of a lsc, semi-convex function ϕ on a metric space (X, d) via the discrete variational scheme employing the *Moreau–Yosida approximation*:

$$\phi_\tau(x) := \inf_{z\in X} \left\{ \phi(z) + \frac{d^2(x, z)}{2\tau} \right\}, \quad x \in X,\ \tau > 0. \tag{1.30}$$

A point x_τ attaining the above infimum is considered as an approximation of the point $\xi(\tau)$ on the gradient curve ξ of ϕ with $\xi(0) = x$. Here ϕ is said to be *semi-convex* if it is λ-*convex* for some $\lambda \in \mathbb{R}$ meaning that equation 1.29 holds along geodesics $\gamma : [0,1] \mapsto X$. Since the approximation scheme is based on the distance function, finer properties of the distance function provide finer or more subtle analysis of gradient flows. In Jost (1995)[56], Mayer (1998) [77] (with applications to harmonic maps) and Ambrosio, Gigli and Savaré (2005) [4] (with applications to the Wasserstein spaces), gradient flows in CAT(0)-*spaces* (non-positively curved metric spaces) are well investigated; see also Bačák's recent book [13]. There the 2-convexity of the squared distance function, that is indeed the definition of CAT(0)-spaces, played essential roles.

In Ohta and Pálfia (2017) [90] is generalized the theory to CAT(1)-*spaces* (metric spaces of sectional curvature ≤ 1), where the distance functions are only semi-convex. CAT(1)-spaces can have a more complicated global structure than CAT(0)-spaces. For instance, all CAT(0)-spaces are contractible while CAT(1)-spaces may not be. Also metric trees are CAT(0)-spaces, but more generally metric graphs are CAT(1) locally and are only CAT(0) if they are trees.

It is known that the direct application of the techniques of CAT(0)-spaces to CAT(1)-spaces does not work. The point is that the K-convexity of the squared distance function with $K < 2$ holds true on some non-Hilbert Banach spaces, on those the behavior of gradient flows is much less understood. It is circumvented this difficulty by introducing the notion of "commutativity" representing a "Riemannian nature" of the space. Precisely, the key ingredients of the analysis in [90] are the following:

(A) The *commutativity* (cf. [90]):

$$\lim_{s\downarrow 0} \frac{d^2(\gamma(s), z) - d^2(x, z)}{s} = \lim_{t\downarrow 0} \frac{d^2(\eta(t), y) - d^2(x, y)}{t} \tag{1.31}$$

for geodesics γ and η with $x = \gamma(0) = \eta(0)$, $\gamma(1) = y$ and $\eta(1) = z$;

(B) The semi-convexity of the squared distance function (see Lemma 1.4.3).

One, if needed, instead of geodesic curves, can more generally consider some other family of curves along which (A) and (B) hold, as it was essentially used to study the Wasserstein spaces in Ambrosio, Gigli and Savaré (2005) literature [4]. This approach, reinforcing the semi-convexity with the Riemannian nature of the space, seems to be of independent interest and is in a similar spirit to Savaré's work [99] based on the semi-concavity of distance functions and the *local angle condition*, those properties fit the study of spaces with lower curvature bounds. The semi-convexity and semi-concavity are usually studied in the context of Banach space theory and Finsler geometry, see [15], [87], [88]. In the CAT(0)-setting, the commutativity follows from the 2-convexity of the squared distance function and the role of the commutativity is implicit. The commutativity is not true in non-Riemannian Finsler manifolds (see Remark 1.4.4(b)). Actually, the lack of commutativity is the reason why Ohta and Sturm (2012) introduced the notion of *skew-convexity* in [91] to study the contraction property of gradient flows in Finsler manifolds. In contrast, on the Wasserstein space over a Riemannian manifold, exists a sort of Riemannian structure, but the convexity of the squared distance function fails. See [46, §B] for a connection between (1.31) and the *infinitesimal Hilbertianity*, which is an "almost everywhere" notion of Riemannian nature.

Let (X, d) be a complete metric space. A curve $\gamma : [0,1] \mapsto X$ is called a *geodesic* if it is locally minimizing and of constant speed. The γ a *minimal* geodesic if it is globally minimizing, namely $d(\gamma(s), \gamma(t)) = |s-t| d(\gamma(0), \gamma(1))$ for all $s, t \in [0,1]$. It can be said that (X, d) is *geodesic* if any two points $x, y \in X$ admit a minimal geodesic between them.

As a potential function, Ohta and Pálfia (2017) consider a lsc function $\phi : X \mapsto (-\infty, \infty]$ such that

$$D(\phi) := X \setminus \phi^{-1}(\infty) \neq \emptyset.$$

Given $x \in X$ and $\tau > 0$, Ohta and Pálfia (2017) consider the *Moreau–Yosida approximation* $\phi_\tau(x)$ in (1.30) and set

$$J_\tau^\phi(x) := \left\{ z \in X \;\middle|\; \phi(z) + \frac{d^2(x,z)}{2\tau} = \phi_\tau(x) \right\}$$

which is called the *Moreau–Yosida resolvent.* For $x \in D(\phi)$ and $z \in J_\tau^\phi(x)$ (if $J_\tau^\phi(x) \neq \emptyset$), it is straightforward from $\phi(z) + \frac{d^2(x,z)}{2\tau} \leq \phi(x)$ that $\phi(z) \leq \phi(x)$ and $d^2(x,z) \leq 2\tau\{\phi(x) - \phi(z)\}$. In this aspect, Ohta and Pálfia (2017) considered two auxiliary conditions on ϕ:

Assumption 1. (1) There exists $\tau_*(\phi) \in (0, \infty]$ such that $\phi_\tau(x) > -\infty$ and $J_\tau^\phi(x) \neq \emptyset$ for all $x \in X$ and $\tau \in (0, \tau_*(\phi))$ (*coercivity*).

(2) For any $Q \in \mathbb{R}$, bounded subsets of the sub-level set $\{x \in X \,|\, \phi(x) \leq Q\}$ are relatively compact in X (*compactness*).

Ohta and Pálfia (2017) observed that, if $\phi_{\tau_*}(x_*) > -\infty$ for some $x_* \in X$ and $\tau_* > 0$, then $\phi_\tau(x) > -\infty$ for every $x \in X$ and $\tau \in (0, \tau_*)$ (see [4, Lemma 2.2.1]). Then, if the compactness (2) in Assumption 2 holds, there is $J_\tau^\phi(x) \neq \emptyset$ by the lsc of ϕ (see [4, Corollary 2.2.2]).

Remark 1.4.1. If $\operatorname{diam} X < \infty$ and the compactness (2) in Assumption 2 holds, then the lsc of ϕ implies that every sub-level set $\{x \in X \,|\, \phi(x) \leq Q\}$ is (empty or) compact. Thus ϕ is bounded below and by the assumption that $\tau_*(\phi) = \infty$.

To construct discrete approximations of gradient curves of ϕ, Ohta and Pálfia (2017) consider a *partition* of the interval $[0,\infty)$:

$$\mathscr{P}_\tau = \left\{0 = t_\tau^0 < t_\tau^1 < \cdots\right\}, \qquad \lim_{k\to\infty} t_\tau^k = \infty,$$

and set

$$\tau_k := t_\tau^k - t_\tau^{k-1} \quad \text{for } k \in \mathbb{N}, \qquad |\tau| := \sup_{k\in\mathbb{N}} \tau_k,$$

so from now on τ denotes a partition rather than just a positive constant and τ may be thought of as the sequence of positive constants $\{\tau_k\}$. Ohta and Pálfia (2017) assume $|\tau| < \tau_*(\phi)$. Given an initial point $x_0 \in D(\phi)$,

$$x_\tau^0 := x_0 \text{ and recursively choose } x_\tau^k \in J_{\tau_k}^\phi(x_\tau^{k-1}) \text{ for each } k \in \mathbb{N}. \tag{1.32}$$

Ohta and Pálfia (2017) mentioned $\{x_\tau^k\}_{k\in\mathbb{N}}$ a *discrete solution* of the variational scheme (2.4) associated with the partition $\mathscr{P}_\tau$, which is thought of as a *discrete-time gradient curve* for the potential function ϕ. If one is interested in how such discrete-time curves can be used for optimization and generalization of the law of large numbers into CAT(1)-spaces can further refer to Ohta and Pálfia (2015) [89].

The following *a priori estimates* (see [4, Lemma 3.2.2]) will be useful later on. Ohta and Pálfia (2017) observed that these estimates are easily obtained if ϕ is bounded below.

Lemma 1.4.2 (A priori estimates)**.** Let $\phi : X \mapsto (-\infty,\infty]$ satisfy Assumption 2 (1). Then, for any $x_* \in X$ and $Q, T > 0$, there exists a constant $C = C(x_*, \tau_*(\phi), Q, T) > 0$ such that, if a partition $\mathscr{P}_\tau$ and an associated discrete solution $\{x_\tau^k\}_{k\in\mathbb{N}}$ of (2.4) satisfy

$$\phi(x_0) \le Q, \quad d^2(x_0, x_*) \le Q, \quad t_\tau^N \le T, \quad |\tau| \le \frac{\tau_*(\phi)}{8},$$

then there exists for any $1 \le k \le N$

$$d^2(x_\tau^k, x_*) \le C, \qquad \sum_{l=1}^{k} \frac{d^2(x_\tau^{l-1}, x_\tau^l)}{2\tau_l} \le \phi(x_0) - \phi(x_\tau^k) \le C.$$

In particular, for all $1 \le k \le N$, exists $d^2(x_\tau^{k-1}, x_\tau^k) \le 2C\tau_k$ and

$$d^2(x_0, x_\tau^k) \le \left(\sum_{l=1}^{k} d(x_\tau^{l-1}, x_\tau^l)\right)^2 \le \sum_{l=1}^{k} \frac{d^2(x_\tau^{l-1}, x_\tau^l)}{\tau_l} \cdot \sum_{l=1}^{k} \tau_l \le 2Ct_\tau^k. \tag{1.33}$$

From here on, let $\phi : (-\infty,\infty] \mapsto X$ be λ-*convex* (also called λ-*geodesically convex*) for some $\lambda \in \mathbb{R}$ in the sense that

$$\phi(\gamma(t)) \leq (1-t)\phi(x) + t\phi(y) - \frac{\lambda}{2}(1-t)td^2(x,y) \tag{1.34}$$

for any $x, y \in D(\phi)$ and some minimal geodesic $\gamma : [0,1] \mapsto X$ from x to y. (The existence of a minimal geodesic between two points in $D(\phi)$ is included in the definition, thus in particular $(D(\phi), d)$ is geodesic.) Ohta and Pálfia (2017) observed that the compactness (2) in Assumption 2 implies the coercivity (1) in this case (see [4, Lemma 2.4.8]; and have $\tau_*(\phi) = \infty$ if $\lambda \geq 0$).

Fix an initial point $x_0 \in D(\phi)$. Take a sequence of partitions $\{\mathscr{P}_{\tau_i}\}_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty}|\tau_i| = 0$ and associated discrete solutions $\{x^k_{\tau_i}\}_{k\in\mathbb{N}}$ with $x^0_{\tau_i} = x_0$. Under Assumption 2 (1), by the compactness argument ([4, Proposition 2.2.3]), a subsequence of the interpolated curves

$$\bar{x}_{\tau_i}(0) := x_0, \qquad \bar{x}_{\tau_i}(t) := x^k_{\tau_i} \quad \text{for } t \in (t^{k-1}_{\tau_i}, t^k_{\tau_i}] \tag{1.35}$$

converges to a curve $\xi : [0,\infty) \mapsto D(\phi)$ pointwise in $t \in [0,\infty)$. In general, under the coercivity and λ-convexity of ϕ (but without the compactness), if a curve ξ is obtained as above (called a *generalized minimizing movement*; see [4, Definition 2.0.6]), then it is locally Lipschitz on $(0,\infty)$ and satisfies $\lim_{t\downarrow 0}\xi(t) = x_0$ as well as the *energy dissipation identity*:

$$\phi(\xi(T)) = \phi(\xi(S)) - \frac{1}{2}\int_S^T \{|\dot{\xi}|^2 + |\nabla\phi|^2(\xi)\}\, dt. \tag{1.36}$$

Here

$$|\dot{\xi}|(t) := \lim_{s\to t}\frac{d(\xi(s),\xi(t))}{|t-s|}$$

is the *metric speed* existing at almost all t, and

$$|\nabla\phi|(x) := \limsup_{y\to x}\frac{\max\{\phi(x)-\phi(y),0\}}{d(x,y)}$$

is the (descending) *local slope* (see [4, Theorem 2.4.15]). Ohta and Pálfia (2017) noticed that $|\nabla\phi|$ is lower semi-continuous ([4, Corollary 2.4.10]) and $\lim_{i\to\infty}\phi(\bar{x}_{\tau_i}(t)) = \phi(\xi(t))$ for all $t \geq 0$ ([4, Theorem 2.3.3]). Equation (1.36) can be thought of as a metric formulation of the differential equation $\dot{\xi}(t) = -\nabla\phi(\xi(t))$, thus ξ will be called a *gradient curve* of ϕ starting from x_0. Ohta and Pálfia (2017) observed that one does not have uniqueness of gradient curves in this generality (see [2, Example 4.23] for a simple example in the ℓ^2_∞-space).

Remark 1.4.2. One can relax the scheme by allowing varying initial points: $x^0_{\tau_i} \neq x_0$. Then assuming $x^0_{\tau_i} \to x_0$ and $\phi(x^0_{\tau_i}) \to \phi(x_0)$ yields the same convergence results.

Ohta and Pálfia (2017) referred to (Yu. Burago, D. Burago and S.Ivanov, 2001) [28] for the basics of CAT(1)-spaces and for more general metric geometry.

Given three points $x, y, z \in X$ with $d(x,y) + d(y,z) + d(z,x) < 2\pi$, it can be taken corresponding points $\tilde{x}, \tilde{y}, \tilde{z}$ in the 2-dimensional unit sphere $\mathbb{S}^2$ (uniquely up to rigid motions) such that

$$d_{\mathbb{S}^2}(\tilde{x}, \tilde{y}) = d(x,y), \qquad d_{\mathbb{S}^2}(\tilde{y}, \tilde{z}) = d(y,z), \qquad d_{\mathbb{S}^2}(\tilde{z}, \tilde{x}) = d(z,x).$$

There exists $\triangle\tilde{x}\tilde{y}\tilde{z}$ a *comparison triangle* of $\triangle xyz$ in $\mathbb{S}^2$.

Definition 1.4.1 (CAT(1)-spaces)**.** A geodesic metric space (X, d) is called a CAT(1)-*space* if, for any $x, y, z \in X$ with $d(x,y) + d(y,z) + d(z,x) < 2\pi$ and any minimal geodesic $\gamma : [0,1] \mapsto X$ from y to z, then

$$d(x, \gamma(t)) \leq d_{\mathbb{S}^2}(\tilde{x}, \tilde{\gamma}(t))$$

at all $t \in [0,1]$, where $\triangle\tilde{x}\tilde{y}\tilde{z} \subset \mathbb{S}^2$ is a comparison triangle, that is, a triangle with correspondingly equal side lengths to $\triangle xyz$ and $\tilde{\gamma} : [0,1] \mapsto \mathbb{S}^2$ is the minimal geodesic from $\tilde{y}$ to $\tilde{z}$.

It is readily observed from the definition that each pair of points $x, y \in X$ with $d(x,y) < \pi$ is joined by a unique minimal geodesic. Fundamental examples of CAT(1)-spaces are the following.

Example 1.4.1. (1) A complete, simply connected Riemannian manifold endowed with the Riemannian distance is a CAT(1)-space if and only if its sectional curvature is not greater than 1 everywhere.

(2) All CAT(0)-*spaces*, which are defined similarly to Definition 1.4.1 by replacing $\mathbb{S}^2$ with $\mathbb{R}^2$, are CAT(1)-spaces. Important examples of CAT(0)-spaces are Hadamard manifolds, Hilbert spaces, trees, and Euclidean buildings.

(3) Further examples of CAT(1)-spaces include orbifolds obtained as quotient spaces of CAT(1)-manifolds, and spherical buildings. See in Burago D., Burago Y. and Ivanov S. (2001) [28, §9.1] for more examples.

Remark 1.4.3. For general $\kappa \in \mathbb{R}$, CAT(κ)-spaces are defined in the same manner by employing comparison triangles in the 2-dimensional space form of constant curvature κ. If (X, d) is a CAT(κ)-space, then it is also CAT(κ') for all $\kappa' > \kappa$ and the scaled metric space (X, cd) with $c > 0$ is CAT($c^{-2}\kappa$). Therefore, considering CAT(1)-spaces covers all CAT(κ)-spaces up to scaling.

The properties of CAT(1)-spaces needed in the further discussion are only the semi-convexity of the squared distance function and the commutativity (1.31). The latter is a consequence of the first variation formula. Let us review them.

Lemma 1.4.3 (Semi-convexity of distance functions)**.** Let (X,d) be a CAT(1)-space and take $R \in (0,\pi)$. Then there exists $K = K(R) \in \mathbb{R}$ such that the squared distance function $d^2(x,\cdot)$ is K-convex on the open R-ball $B(x,R)$ for all $x \in X$.

Clearly $K(R) > 0$ if $R < \pi/2$ (see, e.g., [87] for the precise estimate) and $K(R) < 0$ if $R > \pi/2$. It can be defined the *angle* between two geodesics γ and η emanating from the same point $\gamma(0) = \eta(0) = x$ by

$$\angle_x(\gamma,\eta) := \lim_{s,t\downarrow 0} \angle\gamma(s)\widetilde{\gamma(s)}\tilde{x}\eta(t)\widetilde{\eta(t)},$$

where $\angle\gamma(s)\widetilde{\gamma(s)}\tilde{x}\eta(t)\widetilde{\eta(t)}$ is the angle at $\tilde{x}$ of a comparison triangle $\triangle\gamma(s)\widetilde{\gamma(s)}\tilde{x}\eta(t)\widetilde{\eta(t)}$ in $\mathbb{S}^2$. By the definition of the angle, it can be obtained the following (see [28, Theorem 4.5.6, Remark 4.5.12]).

Theorem 1.4.4 (First variation formula)**.** Let $\gamma : [0,1] \mapsto X$ be a geodesic from x to z, and take $y \in X$ with $0 < d(x,y) < \pi$. Then,

$$\lim_{s\downarrow 0} \frac{d(\gamma(s),y) - d(x,y)}{s} = -d(x,z)\cos\angle_x(\gamma,\eta),$$

where $\eta : [0,1] \to X$ is the unique minimal geodesic from x to y.

Let (X,d) be a complete CAT(1)-space and $\phi : X \mapsto (-\infty,\infty]$ satisfy the λ-convexity for some $\lambda \in \mathbb{R}$ and Assumption 2 (1). Note that (1.34) holds along every minimal geodesic since minimal geodesics are unique between points of distance $< \pi$. The next lemma, generalizing [4, Theorem 4.1.2(ii)] to the case where both λ and K can be negative, is a key tool in what follows.

Lemma 1.4.5 (Key lemma, [90])**.** Let $x \in D(\phi)$ and $\tau \in (0, \min\{\pi^2/(2C), \tau_*(\phi)/8\})$ with $C = C(x, \tau_*(\phi), \phi(x), \tau_*(\phi)/8)$ from Lemma 1.4.2. Take $x_\tau \in J_\tau^\phi(x)$. Then for any $y \in D(\phi) \cap B(x_\tau, R - d(x,x_\tau))$ with $R < \pi$ and for $K = K(R)$ as in Lemma 1.4.3,

$$\begin{aligned}
d^2(x_\tau,y) &\le d^2(x,y) - \lambda\tau d^2(x_\tau,y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - \frac{K}{2}d^2(x,x_\tau) \\
&\le d^2(x,y) - \lambda\tau d^2(x_\tau,y) + 2\tau\{\phi(y) - \phi(x_\tau)\} \\
&\quad + \max\{0,-K\}\cdot\tau\{\phi(x) - \phi(x_\tau)\}.
\end{aligned}$$

Remark 1.4.4. (a) Used in the proof of [4, Theorem 4.1.2(ii)] is the direct application of the convexity of ϕ and $d^2(x,\cdot)$ along γ, which implies in the setting

$$\frac{K}{2}d^2(x_\tau,y) \le d^2(x,y) - \lambda\tau d^2(x_\tau,y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - d^2(x,x_\tau).$$

This coincides with the estimate in the literature of Ohta and Pálfia (2017), when $K = 2$.

The commutativity was used to move the coefficient $K/2$ from $d^2(x_\tau, y)$ to $d^2(x, x_\tau)$, then one can efficiently estimate $d^2(x_\tau, y) - d^2(x, y)$.

(b) As already mentioned (see the paragraph including (1.31)), the Riemannian nature of the space (i.e., the angle) is essential in the commutativity (1.31). In fact, on a Finsler manifold (M, F), (1.31) (written using only the distance) implies:

$$g_v(v, w) = g_w(v, w) \qquad \text{for all } v, w \in T_xM \setminus \{0\},\ x \in M.$$

These notations and the basics of Finsler geometry can be found in [91] for instance. Thus for $v \neq \pm w$,

$$\begin{aligned}
&F^2(v + w) + F^2(v - w) = g_{v+w}(v + w, v + w) + g_{v-w}(v - w, v - w) \\
&= g_{v+w}(v, v + w) + g_{v+w}(w, v + w) + g_{v-w}(v, v - w) - g_{v-w}(w, v - w) \\
&= g_v(v, v + w) + g_w(w, v + w) + g_v(v, v - w) - g_w(w, v - w) \\
&= 2g_v(v, v) + 2g_w(w, w) = 2F^2(v) + 2F^2(w).
\end{aligned}$$

This is the parallelogram identity on T_xM and hence F is Riemannian.

The commutativity (1.31) is the essential property connecting the (geodesic) convexity of a function and the contraction property of its gradient flow. On Finsler manifolds, the contraction property is characterized by the *skew-convexity*, which is different from the usual convexity along geodesics (see [91] for details).

1.5 Applications to Gradient Flows

The argument covers two cases (Ohta and Pálfia, 2017). In both cases, (X, d) is complete, $\phi : X \longrightarrow (-\infty, \infty]$ is lsc, λ-convex and $D(\phi) \neq \emptyset$ [90].

1) (X, d) is a CAT(1)-space.

2) (X, d) satisfies the commutativity and the K-convexity of the squared distance function, and ϕ satisfies the coercivity setting.

Under the condition that both $\lambda, K \in \mathbb{R}$ can be negative.

The estimate in Lemma 1.4.5 is worse than the one in [4, Theorem 4.1.2(ii)] because of the generality that K can be less than 2 and even negative. Nonetheless, through the Lemma 1.4.5 is enough to generalize the argumentation in Chapter 4 of (Ambrosio, Gigli and Savaré (2005) [4].

The K-convexity is assumed globally, thus the assertion of Lemma 1.4.5 holds for any $x, y \in D(\phi)$ and $\tau \in (0, \tau_*(\phi))$. The coercivity holds if, for instance, the compactness condition (Assumption 2 (2)) is satisfied. The coercivity is guaranteed by restricting ourselves to balls with radii $\leq R < \pi/4$. In these (convex) balls the squared distance function is K-convex with $K = K(2R) > 0$ from Lemma 1.4.3, and hence $\phi + d^2(x, \cdot)/(2\tau)$ is $(\lambda + K/(2\tau))$-convex. This implies that $J_\tau^\phi(x)$ is nonempty and consists of a single point for $\tau \in (0, -K/(2\lambda))$ even if $\lambda < 0$ (by, for example, [4, Lemma 2.4.8]). By the same reasoning, if $K > 0$, then Assumption 2 (1) is redundant.

To include both cases keeping clarity, Ohta and Pálfia (2017) under the global K-convexity of the squared distance function and Assumption 2 (1). Thus it is implicitly assumed $\operatorname{diam} X < \pi/2$. This costs no generality for the construction of gradient curves since it is a local problem. Ohta and Pálfia (2017) explained how to extend the properties of the gradient flow to the case of $\operatorname{diam} X \geq \pi/2$.

Given an initial point $x_0 \in D(\phi)$ and a partition $\mathscr{P}_\tau$ with $|\tau| < \tau_*(\phi)$, Ohta and Pálfia (2017) fixed a discrete solution $\{x_\tau^k\}_{k\in\mathbb{N}}$ of (2.4). Let us also take a point $y \in X$. Similarly to Chapter 4 of [4], by the interpolation of the discrete data x_τ^k, $d(x_\tau^k, y)$ and $\phi(x_\tau^k)$ as follows (recall (1.35)):

For $t \in (t_\tau^{k-1}, t_\tau^k]$, $k \in \mathbb{N}$, define

$$\begin{aligned}
\bar{x}_\tau(t) &:= x_\tau^k \in J_{\tau_k}^\phi(x_\tau^{k-1}) \quad (\bar{x}_\tau(0) := x_0), \\
\bar{d}_\tau(t; y) &:= \left\{ d^2(x_\tau^{k-1}, y) + \frac{t - t_\tau^{k-1}}{\tau_k}\{d^2(x_\tau^k, y) - d^2(x_\tau^{k-1}, y)\} \right\}^{1/2}, \\
\bar{\phi}_\tau(t) &:= \phi(x_\tau^{k-1}) + \frac{t - t_\tau^{k-1}}{\tau_k}\{\phi(x_\tau^k) - \phi(x_\tau^{k-1})\}.
\end{aligned}$$

Recall that $\tau_k = t_\tau^k - t_\tau^{k-1}$ and note that $\bar{\phi}_\tau$ is non-increasing.

1.6 Properties of gradient curves

Theorem 1.6.1 (Unique limits of discrete solutions, [90])**.** Fix an initial point $x_0 \in D(\phi)$ and consider discrete solutions $\{x^k_{\tau_i}\}_{k\in\mathbb{N}}$ with $x^k_{\tau_i} = x_0$ associated with a sequence of partitions $\{\mathscr{P}_{\tau_i}\}_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty}|\tau_i|= 0$. Then, the interpolated curve $\bar{x}_{\tau_i}$: $[0,\infty) \mapsto X$ converges to a curve $\xi : [0,\infty) \mapsto X$ with $\xi(0) = x_0$ as $i \to \infty$ uniformly on each bounded interval $[0,T]$. In particular, the limit curve ξ is independent of the choice of the sequence of partitions nor discrete solutions.

By virtue of the uniqueness, it can be defined the *gradient flow* operator:

$$\mathcal{G} : [0,\infty) \times D(\phi) \longrightarrow D(\phi) \tag{1.37}$$

by $\mathcal{G}(t,x_0) := \xi(t)$, where $\xi : [0,\infty) \longrightarrow X$ is the unique gradient curve with $\xi(0) = x_0$ given by Theorem 2.2.11. Then the semi-group property:

$$\mathcal{G}(t,\mathcal{G}(s,x_0)) = \mathcal{G}(s+t,x_0) \quad \text{for all } s,t \geq 0$$

also follows from the uniqueness of gradient curves.

Definition 1.6.1 (Gradient curve $\mathcal{G}(t,x_0) := \xi(t)$ in [73])**.** Gradient flow (or steepest descent curve) is a smooth curve $x : \mathbb{R} \to X$ such that, x(t) is a position at time t and f is the function to minimize, indicated as the negative gradient vector in the form of a differential equation, where the function is decreasing the quickest. It is applied in convergence analysis, since gradient flow gradually minimizes $\xi(x(t))$, and in many optimization methods related to it. Then

$$x'(t) = -\nabla f(x(t)) \tag{1.38}$$

Therefore, $\xi(t) := \mathcal{G}(t,x_0)$ and $\zeta(t) := \mathcal{G}(t,y_0)$ as a point $x_0 \in D(\phi)$ satisfies $|\nabla\phi|(x_0) = 0$ iff $\mathcal{G}(t,x_0) = x_0$ for all $t > 0$ and a point $y_0 \in D(\phi)$ satisfies $|\nabla\phi|(y_0) = 0$ iff $\mathcal{G}(t,y_0) = x_0$ for all $t > 0$.

It can be obtained the following (rough) error estimate based on the proof of Theorem 2.2.11 (compare this with [4, Theorem 4.0.9]). This error estimate below along with others discussed in this section was used later in the paper (Gál and Pálfia, 2024) [44] to prove convergence of gradient curves under Mosco convergence [34] of semi-convex functions in CAT(1)-spaces generalizing the existing results of the CAT(0) setting due to Jost [58], Kuwae-Shioya [66] and Bačák [12].

Corollary 1.6.2 (An error estimate , [90])**.** Let $\lambda \leq 0$ and $|\tau|< \tau_*(\phi)$, fix $x_0 \in D(\phi)$, and put $\xi(t) := \mathcal{G}(t, x_0)$. Then,

$$\bar{d}_\tau^2(t; \xi(t)) \le e^{-2\lambda t} \left(\sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|} \right)^2 |\tau| \{\phi(x_0) - \phi(\bar{x}_\tau(t))\}$$

for all $t > 0$, where $K' := \min\{0, K\}$.

Theorem 1.6.3 (Contraction property, [90])**.** Take $x_0, y_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$ and $\zeta(t) := \mathcal{G}(t, y_0)$. Then exists, for any $t > 0$,

$$d(\xi(t), \zeta(t)) \le e^{-\lambda t} d(x_0, y_0). \tag{1.39}$$

Theorem 1.6.4 (Evolution variational inequality, [90])**.** Take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then,

$$\limsup_{\varepsilon \downarrow 0} \frac{d^2(\xi(t+\varepsilon), y) - d^2(\xi(t), y)}{2\varepsilon} + \frac{\lambda}{2} d^2(\xi(t), y) + \phi(\xi(t)) \le \phi(y) \tag{1.40}$$

for all $y \in D(\phi)$ and $t > 0$. In particular,

$$\frac{1}{2} \frac{d}{dt}[d^2(\xi(t), y)] + \frac{\lambda}{2} d^2(\xi(t), y) + \phi(\xi(t)) \le \phi(y)$$

for all $y \in D(\phi)$ and almost all $t > 0$.

Following the argumentation in the literature of *Gradient flows on non-positively curved metric spaces and harmonic maps* (Mayer, 1998) [77] (on CAT(0)-spaces), Ohta and Pálfia (2017) study stationary points and the large-time behaviour of the gradient flow $\mathcal{G}$. Since the fundamental properties of the flow, for establishing those the CAT(0)-property is used in [77, Section 1], is already at hand, they followed the line of [77, Section 2] in detail. They began with a consequence of the evolution variational inequality (EVI) in Theorem 1.6.4 corresponding to Mayer (1998) [77, Lemma 2.8].

Lemma 1.6.5 ([90])**.** Take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then,

$$d^2(\xi(T), y) \le e^{-\lambda T} d^2(x_0, y) + 2e^{-\lambda T} \int_0^T e^{\lambda t} \{\phi(y) - \phi(\xi(t))\}\, dt$$

for all $T > 0$ and $y \in D(\phi)$. In particular, it exists

$$d^2(\xi(T), y) \le e^{-\lambda T} d^2(x_0, y) - \frac{2(1 - e^{-\lambda T})}{\lambda} \{\phi(\xi(T)) - \phi(y)\},$$

where $(1 - e^{-\lambda T})/\lambda := T$, when $\lambda = 0$.

By the above lemma, one can show a characterization of stationary points of the flow $\mathcal{G}$ in terms of the local slope $|\nabla \phi|$.

Theorem 1.6.6 ([90])**.** A point $x_0 \in D(\phi)$ satisfies $|\nabla\phi|(x_0) = 0$ if and only if $\mathcal{G}(t, x_0) = x_0$ for all $t > 0$.

It is natural to expect that, if $\phi \circ \xi$ does not diverge to $-\infty$, then $|\nabla\phi|\circ\xi$ tends to 0. This is indeed the case as follows.

Lemma 1.6.7 ([90])**.** Assume $\lambda \leq 0$, take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then,

$$|\nabla\phi|(\xi(T)) - |\nabla\phi|(\xi(S)) \leq -\sqrt{2}\lambda \int_S^T |\nabla\phi|\circ\xi \, dt \tag{1.41}$$

for all $0 < S < T$.

Theorem 1.6.8 (Large time behavior, [90])**.** Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume $\lim_{t\to\infty} \phi(\xi(t)) > -\infty$. Then the $\lim_{t\to\infty}|\nabla\phi|(\xi(t)) = 0$.

The following corollary is immediate, see [77, Corollary 2.31].

Corollary 1.6.9 ([90])**.** Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume that there is a sequence $\{t_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} t_n = \infty$ and $\{\xi(t_n)\}_{n\in\mathbb{N}}$ converges to a point $\bar{x}$. Then $\bar{x}$ is a stationary point of ϕ (in the sense of Theorem 1.6.6) and $\lim_{t\to\infty} \phi(\xi(t)) = \phi(\bar{x})$.

In general, $\lim_{t\to\infty}|\nabla\phi|(\xi(t)) = 0$ does not imply the convergence to a stationary point. One needs some compactness condition to find a stationary point, see for instance [77, Theorem 2.32].

Now all the results in this section are generalized in [90] to complete CAT(1)-spaces (X, d) with $\operatorname{diam} X \geq \pi/2$. First of all, given $x_0 \in D(\phi)$, one can restrict the construction of the gradient curve in, say, the open ball $B(x_0, \pi/6)$. Since $B(x_0, \pi/6)$ is (geodesically) convex, the squared distance function in this ball is K-convex with $K = K(\pi/3) > 0$ from Lemma 1.4.3, and it can be obtained the gradient curve ξ with $\xi(0) = x_0$. Once $\xi(t)$ hits the boundary $\partial B(x_0, \pi/6)$ at $t = t_1$, it can be restarted the construction in $B(\xi(t_1), \pi/6)$. Ohta and Pálfia (2017) remark that $t_1 \geq (\pi/6)^2/(2C)$ by (2.5). Iterating this procedure gives the gradient curve $\xi : [0, \infty) \longrightarrow X$. The contraction property and the evolution variational inequality are globalized also in a standard way in [90] using similar ideas. Then Theorems 1.6.6, 1.6.8 also hold true since they are based only on the evolution variational inequality (EVI).

CHAPTER 2

Mosco convergence of semi-convex functions in CAT(1)-space

Convergence of semi-convex functions in the sense of Mosco can be used to describe a new theoretical characterization of a Cartan-Alexandrov-Topogonov (CAT) metric space with upper curvature bound [1]. *Mosco convergence* means that the convergence of non-linear functionals applies weak and strong topologies in the vector space. In my view, we can speak about the term of *Mosco distance*, similarly as the Gromov-Hausdorff convergence considering a Gromov-Hausdorff distance. Particularly, the Mosco distance and Mosco convergent sequences refer to the lower semi-continuous (lsc) semi-convex functions and their convergence of Moreau-Yosida resolvents (Gál and Pálfia, 2024). Apparently, these resolvents take values along the metric space [44]. The properties of Mosco convergence in the complete CAT(0) or Hadamard space was first established by [66] K. Kuwae and T. Shioya (2008) and later by [12] M. Bačak (2015). The latter book provides a comprehensive overview of a properties related to CAT(0)-space, which literature by M. Bačak (2015) was read before the investigation into the CAT(1)-space. Applications in this chapter additionally incorporates the a priori estimation of [90] Shin-ichi Ohta and M. Pálfia (2015) in CAT(1)-spaces on the theory of gradient flows.

The Mosco convergence implies convergence of resolvents, which in turn entails the convergence of gradient flows for lower-semi-continuous (lsc) semi-convex functions.

[1]See as a reference: Gál, H., Pálfia, M. *Convergence of semi-convex functions in CAT(1)-spaces. Calculus of Variations and Partial Differential Equations.* 63, 213 (2024), pp. 1-20. https://doi.org/10.1007/s00526-024-02823-4.

2.0 Motivation and the research question

The **research question** refers to the fact that under condition of a CAT(1)-space with $diam(x) < \pi$, the sequence of λ-convex lsc and L-Lipschitz function $f_n : X \to (-\infty, \infty]$ Mosco converging in the weak sense to $f : X \to (-\infty, \infty]$, in which way we can approximate the gradient curve? How we can get the convergence of a gradient flow? The **methodology** relates to the techniques, which utilize weak convergence in the CAT(1)-spaces and also cover asymptotic relations of sequences of such spaces introduced by K. Kuwae and T. Shioya (2008), including Gromov-Hausdorff limits. The application is associated with the following areas: topology of a sphere, the Geometric Group Theory (GGT), trigonometry (Alexandrov angle and the Law of cosine), Functional Analysis, Optimization (Ekeland's variational principle and Moreau-Yosida approximation), gradient flow, Evolution Variational Inequality (EVI).

2.1 Weak convergence in CAT(1)-spaces

We begin with recalling the concept of *weak convergence* in a CAT(1) space [57] (Jost, 1994), see also Definition 27. in [106] (Yokota, 2016). In the definition, we make use of $P_Z(x)$, which denotes the unique nearest point projection of the point $x \in X$ onto a closed π-convex set $Z \subseteq X$ with $d(Z, x) := \inf_{z \in Z} d(z, x) < \pi/2$. Under these conditions, $P_Z(x)$ is a retraction with the obtuse-angle property, see for example Exercise 2.6(1) in [27] (Bridson and Haefliger, 1999).

Definition 2.1.1 (Weak convergence)**.** Let (X, d) be a CAT(1) space. A net $x_i \in X$ included in $B_x(\pi/2)$ with $x \in X$ *weakly converges* to the point x, that is $x_i \xrightarrow{w} x$, if for any π-convex geodesic segment $\gamma \subseteq X$ with $\gamma(0) = x$, $P_\gamma(x_i)$ strongly converges to x.

In order to investigate into weak convergence in CAT(1)-space, we need prior to look into the strong convergence in CAT(0)-space. The Mosco convergence of a sequence of functions implies strong convergence of the resolvents and semigroups (Bačák, 2015) [12]. Strong convergence is defined as the condition, where the sequence of function x_i converges to x (Moore, 2007) [80].

Definition 2.1.2 (Strong convergence)**.** Lemma 3.15 and Equation 21. in [12] together provide the strong convergence. Let $x_i \in X$ be a bounded sequence and $y_i \in X$. Assume $x_i \xrightarrow{w} x \in X$ and $y_i \to y \in X$. Then,

1. $d(x_i, y_i) \leq \liminf_{i \to \infty} d(x_i, y_i)$;
2. $d(x_i, y_i) \to d(x, y) iff x_i \to x$.

Onwards, let's see the strong convergence of resolvents $J_\lambda^n \to J_\lambda$ as $n \to \infty$:

Proof. Now, we are in position to prove $\lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x)$. Suppose on the contrary that there exists a subsequence $j_k \in J_\tau^{f_{n_k}}(x)$ such that $d(j_k, J_\tau^f(x))$, does not converge to 0. Since $\operatorname{diam}(X) < \pi$, there exists a subsequence of $\{j_k\}_{k\in\mathbb{N}}$ still denoted by j_k with a weak limit $c \in X$. Since, $f_n \to f$ in the sense of Mosco, there exists a sequence $y_n \to J_\tau^f(x)$ and $f_n(y_n) \to f(J_\tau^f(x))$. Then, using the weak lsc property of $d^2(x,\cdot)$ we get:

$$\begin{aligned}
\limsup_{k\to\infty}(f_{n_k})_\tau(x) &\leq \limsup_{k\to\infty} f_{n_k}(y_{n_k}) + \frac{1}{2\tau}d^2(x, y_{n_k}) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau}d^2(x, J_\tau^f(x)) \\
&\leq f(c) + \frac{1}{2\tau}d^2(x,c) \\
&\leq \liminf_{k\to\infty} f_{n_k}(j_k) + \frac{1}{2\tau}d^2(x, j_k) \\
&= \liminf_{k\to\infty} f_{n_k}(J_\tau^{f_{n_k}}(x)) + \frac{1}{2\tau}d^2(x, J_\tau^{f_{n_k}}(x)) \\
&= \liminf_{k\to\infty}(f_{n_k})_\tau(x)
\end{aligned} \tag{2.1}$$

which yields $c \in J_\tau^f(x)$. Furthermore, using (M1) and (M2) with $z_n \to c$ and $f_n(z_n) \to f(c)$, we have:

$$\begin{aligned}
\limsup_{k\to\infty} \frac{1}{2\tau}d^2(x, j_k) &\leq \limsup_{k\to\infty} -f_{n_k}(j_k) + f_{n_k}(z_{n_k}) + \frac{1}{2\tau}d^2(x, z_{n_k}) \\
&= -\liminf_{k\to\infty} f_{n_k}(j_k) + f(c) + \frac{1}{2\tau}d^2(x,c) \\
&\leq -f(c) + f(c) + \frac{1}{2\tau}d^2(x,c) \\
&= \frac{1}{2\tau}d^2(x,c) \\
&\leq \liminf_{k\to\infty} \frac{1}{2\tau}d^2(x, j_k)
\end{aligned}$$

which, together with (2.3) and Lemma 2.1.1 proves the strong convergence:

$$j_k \to c$$

which, contradicts the assumption that $d(j_k, J_\tau^f(x))$ as $\kappa \to \infty$. Then, using (2.1), we get:

$$\begin{aligned}
\lim_{n\to\infty}(f_n)_\tau(x) &= \lim_{n\to\infty} f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau}d^2(x, J_\tau^{f_n}(x)) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau}d^2(x, J_\tau^f(x)) = f_\tau(x)
\end{aligned}$$

which completes the proof of the theorem. □

The following result characterizes weak convergence, this lemma can be found in [12] (Bačák, 2015) in the case of CAT(0)-spaces. In the CAT(1)-case the price to be paid in the formulation is boundedness implied by the required convexity of geodesic lines.

Lemma 2.1.1. Let (X, d) be a CAT(1) space. Let $x_n, x \in X$ such that $d(x_n, x) < \pi/2$ for all $n \in \mathbb{N}$. Then $x_n \to x$ if and only if $x_n \xrightarrow{w} x$ and $d(x_n, y) \to d(x, y)$ for some $y \in X$ such that $d(x, y) < \pi/2$.

Proof. The " $\Rightarrow$ " is trivial, so we only need to prove the " $\Leftarrow$ " implication. Let $x_n \xrightarrow{w} x$ and $d(x_n, y) \to d(x, y)$ for some $y \in X$, such that $d(x, y), d(x_n, x) < \pi/2$. We can assume that $d(y, x) > 0$, otherwise $d(x_n, x) \to d(x, x) = 0$ implying already that $x_n \to x$. Then, by Exercise 2.6(1) [27] we have that the Alexandrov-angle $\angle_{P_{[x,y]}(x_n)}(x_n, y) \geq \pi/2$, which is the *obtuse-angle property*, and through the spherical law of cosines we have:

$$\begin{aligned}\cos d(x_n, y) &\leq \cos d(x_n, P_{[x,y]}(x_n)) \cos d(y, P_{[x,y]}(x_n)) \\ &\quad + \sin d(x_n, P_{[x,y]}(x_n)) \sin d(y, P_{[x,y]}(x_n)) \cos \angle_{P_{[x,y]}(x_n)}(x_n, y) \\ &\leq \cos d(x_n, P_{[x,y]}(x_n)) \cos d(y, P_{[x,y]}(x_n)).\end{aligned}$$

Now, the weak convergence implies the convergence of projections: $P_{[x,y]}(x_n) \to x$. Moreover, since $P_{[x,y]}(x_n) \to x$, we may assume $P_{[x,y]}(x_n) \neq y$. Thus, the assumption $d(x_n, y) \to d(x, y)$ with the above estimate implies:

$$\cos d(x_n, P_{[x,y]}(x_n)) \to 1, \tag{2.2}$$

thus, $x_n \to x$ as well. □

We also have the following fact, which ensures the small enough bounded sets are weakly sequentially compact, and hence has the Banach-Alaoglu property.

Lemma 2.1.2 (cf. Lemma 28 in [106] (Yokota, 2016))**.** Let (X, d) be a complete CAT(1)-space. Any sequence $(p_n)_{n\in\mathbb{N}}$ of points in X included in a π-convex set has a subsequence, which converges weakly to a point in X.

We will make use of the following result later.

Lemma 2.1.3. Let (X, d) be a CAT(1) space with $\operatorname{diam}(X) < \pi$. Then, if $(C_i)_{i\in I}$ is a non-increasing family of closed convex sets in X for an index set I, we have $\cap_{i\in I} C_i \neq \emptyset$.

Proof. See for example [62, Theorem 2.5. (Kell, 2014)] □

2.2 Ekeland principle and Mosco continuity

Let (X, d) be a CAT(1)-space in this section. We denote the extended real numbers by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$.

Lemma 2.2.1 (Ohta's lemma cf. Lemma 5 in [106] (Yokota, 2016))**.** Any geodesic $\gamma : [0, 1] \mapsto X$ with $\gamma(0), \gamma(1) \in \overline{B}_z((\pi - \epsilon)/2)$, for an $\epsilon > 0$, satisfies

$$d^2(\gamma(t), z) \leq (1 - t)d^2(\gamma(0), z) + td^2(\gamma(1), z) - \frac{\kappa}{2}t(1 - t)d^2(\gamma(0), \gamma(1))$$

for any $t \in [0, 1]$ and $\kappa = (\pi - \epsilon)\tan(\epsilon/2)$.

Lemma 2.2.2. Let $\mathrm{diam}(X) < \pi$. Let $f : X \mapsto (-\infty, \infty]$ be a convex lsc function. Then f is bounded below.

Proof. Assume on the contrary that $\inf_X f = -\infty$. Then, the sub-level sets $X_k := \{x \in X : f(x) \leq -k\}$ for $k \in \mathbb{N}$ are nonempty and are bounded, closed and convex. The sequence $\{X_k\}_{k\in\mathbb{N}}$ is non-increasing, therefore by Lemma 2.1.3 they have non-empty intersection containing a point y. Then, by the construction of $f(y) = -\infty$, which actually represents a contradiction. □

Theorem 2.2.3 (Weakly closed convex set in *Theorem 3.5.* (Kell, 2014))**.** Let $\mathrm{diam}(X) < \pi$. Then, the closed convex sets are weakly closed [62].

Lemma 2.2.4 (Weakly lsc function in *Proposition 3.8.* (Kell, 2014))**.** Let $\mathrm{diam}(X) < \pi$. Let $f : X \to (-\infty, \infty]$ be a quasi-convex lsc function. Then f is weakly lsc. In particular $x \to d^2(a, x)$ is weakly lsc on $B_a(\pi/2)$ [62].

Proof. Assume on the contrary that

$$\liminf_{x \xrightarrow{w} x_0} f(x) < f(x_0).$$

This means that, there exists a subsequence x_{n_k}, and index $k_0 \in \mathbb{N}$, and a $\delta > 0$, such that, $f(x_{n_k}) < f(x_0) - \delta$ for all $k > k_0$. Using lsc and quasi-convexity of f, we get

$$f(y) \leq f(x_0) - \delta$$

for all $y \in \overline{\mathrm{co}}\{x_{n_k} : k > k_0\}$. This contradicts $x_0 \in \overline{\mathrm{co}}\{x_{n_k} : k > k_0\}$ by Lemma 3.2. [62](Kell, 2014).

Fact 4 in [106] (Yokota, 2016) states about the local uniform convexity, referring on Ohta's Lemma 3.1 in [87] of diam $< \frac{\pi}{2}$ as any $\kappa, \epsilon, r > 0$, there is $d_\kappa(\epsilon; r) > 0$ for CAT(1)-space. □

We need the following result of Kendall, which assures the existence of jointly convex functions on CAT(1)-spaces with convex geometry.

Theorem 2.2.5 (Yokota's Theorem A in [106] (Yokota, 2016))**.** Let $\operatorname{diam}(X) < \pi$. There exists a jointly κ-convex lsc function $\Phi : X \times X \to [0, \infty)$ vanishing on the diagonal set $\Delta := \{(x, x) | x \in X\}$ for some $\kappa > 0$.

Lemma 2.2.6 (Ekeland principle, cf. [38] (Ekeland, 1979))**.** Given $x_0 \in X$ and a lsc function $f : X \mapsto (-\infty, \infty]$ that is bounded below, there exist $\alpha, \beta \geq 0$, such that

$$f(x) \geq -\alpha d(x, x_0) - \beta$$

for all $x \in X$.

Definition 2.2.1 (Mosco convergence)**.** A sequence of lsc functions $f_n : X \mapsto \overline{\mathbb{R}}$ said to converge to $f : X \mapsto \overline{\mathbb{R}}$ in the sense of Mosco if, for any $x \in X$, we have

(M1) $f(x) \leq \liminf_{n\to\infty} f_n(x_n)$ whenever $x_n \xrightarrow{w} x$,

(M2) there exists an $(y_n) \subseteq X$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$.

Proposition 2.2.7 (Ekeland principle for a Mosco convergent sequence of lsc semi–convex functions, bounded case)**.** Let X be a CAT(1)-space with $\operatorname{diam}(X) < \pi$. Given $x_0 \in X$ and a sequence of lsc λ-convex functions $f_n : X \to (-\infty, \infty]$ that is Mosco converging to $f : X \to (-\infty, \infty]$, there exist $\alpha, \beta \geq 0$ such that

$$f_n(x) \geq -\alpha d(x, x_0) - \beta$$

for all $x \in X$ and $n \in \mathbb{N}$.

Proof. Assume that the assertion is false; that is, for any $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ and $x_k \in X$, such that:

$$f_{n_k}(x_k) + k[d(x_k, x_0) + 1] < 0.$$

First, assume that $n_k \to \infty$ as $k \to \infty$. Since, we have $\operatorname{diam}(X) < \pi$, by Lemma 2.1.2 we can select a weakly convergent subsequence from x_k still denoted by x_k with weak limit denoted by $\overline{x} \in X$. Then, by the Mosco convergence of f_n we have

$$\begin{aligned} f(\overline{x}) \leq \liminf_{k\to\infty} f_{n_k}(x_k) &\leq \liminf_{k\to\infty} -k[d(x_k, x_0) + 1] \\ &= -\limsup_{k\to\infty} k[d(x_k, x_0) + 1] \\ &= -\infty \end{aligned}$$

which is a contradiction.

What is left to check, it is the case, when n_k remains bounded; that is, after some index $k_0 \in \mathbb{N}$ we have $n_k = n_{k_0}$ for all $k \geq k_0$. Let x_0 denote the circumcenter of X, which exists by Lemma 3.3 [16] (Balser and Lytchak, 2005) and satisfies $d(x_0, x) < \frac{\pi}{2}$ uniformly for all $x \in X$. Consider the lsc function $g(x) := f_{n_{k_0}}(x) + cd^2(x_0, x)$. Since $f_{n_{k_0}}$ is λ-convex, we claim that there exists a large $c > 0$ enough, such that $g(x)$ is convex considering it as lsc function. Indeed, using Lemma 2.2.1, we get:

$$\begin{aligned}(1-t)g(\gamma(0)) + tg(\gamma(1)) - g(\gamma(t)) &\geq \left(\frac{\lambda}{2} + c\frac{\kappa}{2}\right) t(1-t)d^2(\gamma(0), \gamma(1)) \\ &\geq 0\end{aligned}$$

if $c > 0$ is large enough. Pick such a $c > 0$. Then by Lemma 2.2.4 $g : X \to (-\infty, \infty]$ is weakly lsc, thus for a weakly convergent subsequence of x_k still denoted by x_k with weak limit $\overline{x} \in X$, we get:

$$\begin{aligned} g(\overline{x}) &\leq \liminf_{k\to\infty} f_{n_{k_0}}(x_k) + cd^2(x_0, x_k) \\ &\leq \liminf_{k\to\infty} -k[d(x_k, x_0) + 1] + cd^2(x_0, x_k) \\ &\leq -\limsup_{k\to\infty} k[d(x_k, x_0) + 1] + \liminf_{k\to\infty} cd^2(x_0, x_k) \\ &\leq -\limsup_{k\to\infty} k[d(x_k, x_0) + 1] + \text{const.} \\ &= -\infty \end{aligned}$$

which is a contradiction. □

For a potential function, we always consider a lower semi-continuous function $\phi : X \to \overline{\mathbb{R}}$. We recall that the *effective domain* of ϕ is defined as non-empty. Given $x \in X$ and $\tau > 0$, we define the *Moreau–Yosida approximation*:

$$\phi_\tau(x) := \inf_{z\in X} \left\{ \phi(z) + \frac{d^2(x,z)}{2\tau} \right\}$$

and the *Moreau–Yosida resolvent* set:

$$J_\tau^\phi(x) := \left\{ z \in X \,\middle|\, \phi(z) + \frac{d^2(x,z)}{2\tau} = \phi_\tau(x) \right\}.$$

For $x \in D(\phi)$ and $z \in J_\tau^\phi(x)$ (if $J_\tau^\phi(x) \neq \emptyset$), it is straightforward from

$$\phi(z) + \frac{d^2(x,z)}{2\tau} \leq \phi(x)$$

that $\phi(z) \leq \phi(x)$ and $d^2(x,z) \leq 2\tau\{\phi(x) - \phi(z)\}$.

We can consider two kinds of conditions on ϕ:

Assumption 2. 1. There exists $\tau_*(\phi) \in (0, \infty]$, such that $\phi_\tau(x) > -\infty$ and $J_\tau^\phi(x) \neq \emptyset$ for all $x \in X$ and $\tau \in (0, \tau_*(\phi))$ (*coercivity*).

2. For any $Q \in \mathbb{R}$, bounded subsets of the sub-level set $\{x \in X \,|\, \phi(x) \leq Q\}$ are relatively compact in X (*compactness*).

We remark that, if $\phi_{\tau_*}(x_*) > -\infty$ for some $x_* \in X$ and $\tau_* > 0$, then $\phi_\tau(x) > -\infty$ for every $x \in X$ and $\tau \in (0, \tau_*)$ (see [4, Lemma 2.2.1, (Ambrosio, Gigli and Savaré, 2005)]). Then, if the compactness (2) holds, we have $J_\tau^\phi(x) \neq \emptyset$ by the lower semi-continuity of ϕ (see [4, Corollary 2.2.2, (Ambrosio, Gigli and Savaré, 2005)]).

Remark 2.2.1. If $\operatorname{diam} X < \infty$ and the compactness (2) holds, then the lower semi-continuity of ϕ implies that every sub-level set $\{x \in X \,|\, \phi(x) \leq Q\}$ is (empty or) compact. Thus ϕ is bounded below and we can take $\tau_*(\phi) = \infty$, in particular, (1) holds.

The following assures non-emptyness and unicity of $J_\tau^\phi(z)$:

Remark 2.2.2. For λ-convex ϕ by Proposition 3.26 in [64] (Kuwae, 2015) $J_\tau^\phi(z)$ is a unique point in the same neighborhoods of z as in Lemma 2.2.1 for all $K + 2\lambda\tau > 0$, where $K > 0$. The K-convexity of $x \mapsto d^2(z, x)$ holds under $diam(X) < \frac{\pi}{2} - \epsilon$ with $K := (\pi - 2\epsilon)\tan\epsilon$.

The following is one of the main results of this section.

Theorem 2.2.8. Let $\operatorname{diam}(X) < \pi$ and $f_n : X \to (-\infty, \infty]$ a sequence of lsc λ-convex functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Then

$$\lim_{n\to\infty} (f_n)_\tau(x) = f_\tau(x)$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x)$$

for any $\tau > 0$ small enough and $x \in D(f)$.

Further, let's see the proof, which comes from the Proposition 2.2.7.:

Proof. From Proposition 2.2.7, we have for some $\alpha, \beta \geq 0$:

$$f_n(J_\tau^{f_n}(x)) \geq -\alpha d(J_\tau^{f_n}(x), x) - \beta.$$

From the definition of $J_\tau^{f_n}(x)$ and taking $y_n \to x$ as in (M2), we have:

$$f_n(y_n) + \frac{1}{2\tau} d^2(x, y_n) \geq f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau} d^2(x, J_\tau^{f_n}(x))$$

where $y_n \to x$, which combined with the above estimate yields:

$$f_n(y_n) + \frac{1}{2\tau} d^2(x, y_n) + \alpha d(J_\tau^{f_n}(x), x) + \beta \geq \frac{1}{2\tau} d^2(x, J_\tau^{f_n}(x)).$$

That is,

$$0 \geq d^2(x, J_\tau^{f_n}(x)) - 2\tau\alpha d(J_\tau^{f_n}(x), x) - 2\tau(f_n(y_n) + \beta) - d^2(x, y_n).$$

Since, $d^2(x, y_n) \to 0$ and $f_n(y_n) \to f(x) < \infty$ by (M2), for $\tau > 0$ small enough the above forces:

$$d(x, J_\tau^{f_n}(x)) < \epsilon \tag{2.3}$$

for all large enough $n \in \mathbb{N}$, given an arbitrary non-fixed $\epsilon \leq \frac{\pi}{2}$. We know that, ϵ is a variable positive real number that represents an arbitrary small distance in the metric space. □

Remark 2.2.3. If $J_\tau^f(x)$ is not unique in the above Theorem 2.2.8, then it still follows from the same proof that all weak cluster points of $J_\tau^{f_n}(x)$ are in fact strong cluster points and are in $J_\tau^f(x)$.

The following is a variant of the above result using different assumptions.

Theorem 2.2.9. Let $f_n : X \to (-\infty, \infty]$ be a sequence of L-Lipschitz functions that is Mosco converging to $f : X \to \overline{\mathbb{R}}$. Then,

$$\lim_{n\to\infty} (f_n)_\tau(x) = f_\tau(x)$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x)$$

for any small enough $\tau > 0$.

Proof. From the definition of $J_\tau^{f_n}(x)$, we have:

$$f_n(x) \geq f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau} d^2(x, J_\tau^{f_n}(x))$$

which combined with the L-Lipschitz property yields:

$$\begin{aligned} d^2(x, J_\tau^{f_n}(x)) &\leq 2\tau[f_n(x) - f_n(J_\tau^{f_n}(x))], \\ d(x, J_\tau^{f_n}(x)) &\leq 2\tau \frac{f_n(x) - f_n(J_\tau^{f_n}(x))}{d(x, J_\tau^{f_n}(x))} \\ &\leq 2\tau L \end{aligned}$$

From the latter equation related to the small enough $\tau > 0$, we obtain that:

$$d(x, J_\tau^{f_n}(x)) < \epsilon$$

for all large enough $n \in \mathbb{N}$, given an arbitrary non-fixed $\epsilon \leq \frac{\pi}{2}$. We can follow the proof of the previous Theorem 2.2.8 from (2.1) onward to obtain our assertions. $\square$

In the the sequel, we provide another variant of the above theorems. It generalizes Proposition 5.12. in [66] (Kuwae, 2008) in CAT(0)-spaces.

Theorem 2.2.10. Let $f_n : X \to (-\infty, \infty]$ be a uniformly lower bounded sequence of lsc functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Then:

$$\lim_{n\to\infty} (f_n)_\tau(x) = f_\tau(x)$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x)$$

for any small enough $\tau > 0$ and $x \in D(f)$.

Proof. Without loss of generality, we assume the uniform lower bound $f_n, f \geq 0$. Since $f \not\equiv +\infty$, we have $f_\tau(x) < +\infty$ and thus $f(J_\tau^f(x)) < +\infty$. By (M2) there exists an $\{y_n\}_{n\in\mathbb{N}} \subseteq X$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$. Then by definition we have:

$$f_n(y_n) + \frac{1}{2\tau} d^2(x, y_n) \geq f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau} d^2(x, J_\tau^{f_n}(x)),$$

which yield:

$$\begin{aligned} d^2(x, J_\tau^{f_n}(x)) &\leq 2\tau[f_n(y_n) - f_n(J_\tau^{f_n}(x))] + d^2(x, y_n) \\ &\leq 2\tau f_n(y_n) + d^2(x, y_n) \to 2\tau f(x) \end{aligned}$$

as $n \to \infty$. Choosing small enough $\tau > 0$, this eventually yields that:

$$d(x, J_\tau^{f_n}(x)) < \epsilon$$

for all large enough $n \in \mathbb{N}$, given an arbitrary non-fixed $\epsilon \leq \frac{\pi}{2}$. $\square$

To construct discrete approximations of the gradient curves of ϕ, we consider a *partition* of the interval $[0, \infty)$:

$$\mathscr{P}_\tau = \{0 = t_\tau^0 < t_\tau^1 < \cdots\}, \qquad \lim_{k\to\infty} t_\tau^k = \infty,$$

and set

$$\tau_k := t_\tau^k - t_\tau^{k-1} \quad \text{for } k \in \mathbb{N}, \qquad |\tau| := \sup_{k\in\mathbb{N}} \tau_k.$$

We will always assume $|\tau|< \tau_*(\phi)$, which is an order of magnitude distance between partitions. Given an initial point $x_0 \in D(\phi)$,

$$x_\tau^0 := x_0 \text{ and recursively choose } x_\tau^k \in J_{\tau_k}^\phi(x_\tau^{k-1}) \text{ for each } k \in \mathbb{N}. \tag{2.4}$$

We call $\{x_\tau^k\}_{k\in\mathbb{N}}$ a *discrete solution* of the variational scheme (2.4) associated with the partition $\mathscr{P}_\tau$, which is thought of as a *discrete-time gradient curve* for the potential function ϕ. The following a priori estimates (see [4, Lemma 3.2.2, Amrosio, Gigli and Savaré, 2005)]) will be useful in the sequel. We remark that these estimates are easily obtained if ϕ is bounded below.

The *a priori* estimate is already mentioned in Chapter 1., in Lemma 1.8.4.

In particular, for all $1 \le k \le N$, we have $d^2(x_\tau^{k-1}, x_\tau^k) \le 2C\tau_k$ and

$$d^2(x_0, x_\tau^k) \le \left(\sum_{l=1}^k d(x_\tau^{l-1}, x_\tau^l)\right)^2 \le \sum_{l=1}^k \frac{d^2(x_\tau^{l-1}, x_\tau^l)}{\tau_l} \cdot \sum_{l=1}^k \tau_l \le 2Ct_\tau^k. \tag{2.5}$$

The following result ensures the existence of the gradient curve.

Theorem 2.2.11 (Gradient curve, cf. [90], (Ohta and Pálfia, 2017))**.** Fix an initial point $x_0 \in D(\phi)$ and consider discrete solutions $\{x_{\tau_i}^k\}_{k\in\mathbb{N}}$ with $x_{\tau_i}^0 = x_0$ associated with a sequence of partitions $\{\mathscr{P}_{\tau_i}\}_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty}|\tau_i|= 0$. Then, the piecewise interpolated curve $\bar{x}_{\tau_i} : [0,\infty) \mapsto X$ converges to a curve $\xi : [0,\infty) \mapsto X$ with $\xi(0) = x_0$ as $i \to \infty$ uniformly on each bounded interval $[0,T]$. In particular, the limit curve ξ is independent of the choice of the sequence of partitions.

The Error Estimate can be found in Chapter 1., in Corollary 1.8.12.

Thanks to the a priori estimate (Lemma 1.4.2), we have:

$$\max\{\phi(x_0) - \phi(x_{\tau_i}^k), \phi(x_0) - \phi(x_{\tau_j}^l)\} \le C = C(x_0, \tau_*(\phi), \phi(x_0), T).$$

The Theorem of Contraction property is already mentioned in Chapter 1., in Theorem 1.7.2.

The following inequality is proved in Lemma 3.34. in [64] (Kuwae, 2015) for all $K + 2\lambda\tau > 0$, where $K > 0$ is the K-convexity of $x \mapsto d^2(a,x)$. The K-convexity of $x \mapsto d^2(z,x)$ holds under $\mathrm{diam}(X) < \frac{\pi}{2} - \epsilon$ with $K := (\pi - 2\epsilon)\tan\epsilon$. For instance, in Ohta's lemma 2.2.1 we have $K > 0$ on $\overline{B}_a((\pi-\epsilon)/2)$, thus for any $a \in \overline{B}_z((\pi-\epsilon)/4)$ the distance:

$$d(J_\tau^f(x), J_\tau^f(y)) \le 4\frac{2d(x,y) + d(x, J_\tau^f(y)) + d(J_\tau^f(x), y)}{K + 2\lambda\tau} d(x,y) \tag{2.6}$$

also the same states in Lemma 3.34 in [64] (Kuwae, 2015) for any $x, y \in \overline{B}_z((\pi-\epsilon)/4)$ and fixed z. See Lemma 2.3.4.

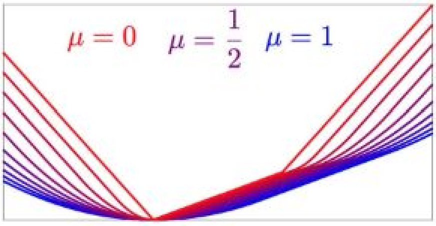

Figure 2.1: Moreau-Yosida regularization. Note: $\tau = \mu$ *Source: Gabriel Peyré, Moreau Yosida envelop theorem, 2022.*

Now, we are in position to prove the main result of this section regarding the continuity of gradient flows under Mosco convergence. The proof also works in CAT(0)-spaces.

Theorem 2.2.12. Let (X, d) be a CAT(1) space. Assume $f_n : X \to (-\infty, \infty]$ is a sequence of lsc semi-convex functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Let $J_\tau^f(x)$ and $J_\tau^{f_n}(x)$ be the corresponding resolvents such that for all small enough $\tau > 0$:

$$J_\tau^{f_n}(x) \to J_\tau^f(x)$$

as established earlier under the extra assumptions imposed by either Theorem 2.2.8 or Theorem 2.2.9 or Theorem 2.2.10. Let $S_t(x) = \mathcal{G}(t, x_0) := \xi(t)$ and $S_t^n(x)$ denote the corresponding gradient flows. Then,

$$\lim_{n\to\infty} S_t^n(x) = S_t(x)$$

for any $x \in D(f)$.

Proof. By (M2) there exists an $y_n \to x$ with $f_n(y_n) \to f(x)$. Then,

$$\begin{aligned} d(S_t(x), S_t^n(x)) &\le d(S_t(x), S_t^n(y_n)) + d(S_t^n(y_n), S_t^n(x)) \\ &\le d(S_t(x), S_t^n(y_n)) + e^{-\lambda t} d(y_n, x). \end{aligned} \tag{2.7}$$

Furthermore,

$$\begin{aligned} d(S_t(x), S_t^n(y_n)) \le d&\left(S_t(x), \left(J_{t/k}^f\right)^k(x)\right) \\ &+ d\left(\left(J_{t/k}^f\right)^k(x), S_t^n(y_n)\right) \\ \le d&\left(S_t(x), \left(J_{t/k}^f\right)^k(x)\right) + d\left(\left(J_{t/k}^f\right)^k(x), \left(J_{t/k}^{f_n}\right)^k(y_n)\right) \\ &+ d\left(\left(J_{t/k}^{f_n}\right)^k(y_n), S_t^n(y_n)\right). \end{aligned} \tag{2.8}$$

To estimate the second term of (2.8) we argue as:

$$\begin{aligned}
&d\left(\left(J^f_{t/k}\right)^k(x), \left(J^{f_n}_{t/k}\right)^k(y_n)\right)\\
&\le d\left(J^f_{t/k}\left(\left(J^f_{t/k}\right)^{k-1}(x)\right), J^{f_n}_{t/k}\left(\left(J^f_{t/k}\right)^{k-1}(x)\right)\right)\\
&\quad + d\left(J^{f_n}_{t/k}\left(\left(J^f_{t/k}\right)^{k-1}(x)\right), J^{f_n}_{t/k}\left(\left(J^{f_n}_{t/k}\right)^{k-1}(x)\right)\right)\\
&\quad + d\left(J^{f_n}_{t/k}\left(\left(J^{f_n}_{t/k}\right)^{k-1}(x)\right), \left(J^{f_n}_{t/k}\right)^k(y_n)\right) \qquad (2.9)\\
&= d\left(J^f_{t/k}\left(\left(J^f_{t/k}\right)^{k-1}(x)\right), J^{f_n}_{t/k}\left(\left(J^f_{t/k}\right)^{k-1}(x)\right)\right)\\
&\quad + d\left(J^{f_n}_{t/k}\left(\left(J^f_{t/k}\right)^{k-1}(x)\right), J^{f_n}_{t/k}\left(\left(J^{f_n}_{t/k}\right)^{k-1}(x)\right)\right)\\
&\quad + d\left(\left(J^{f_n}_{t/k}\right)^k(x), \left(J^{f_n}_{t/k}\right)^k(y_n)\right).
\end{aligned}$$

First, notice that (2.3) is established in all of Theorems 2.2.8,2.2.9,2.2.10. Then, (2.6) establishes a Lipschitz estimate for resolvents in CAT(1) spaces with small enough radius, where (2.3) ensures boundedness of Lipschitz constants, so applying (2.9) recursively for each k we obtain an upper bound, in which the already established pointwise convergence of resolvents assures this second term also goes to 0 for each fixed k as $n \to \infty$. We continue by estimating each of the remaining terms on the right hand side of (2.8). By (1.6.2) with $C_0 := \sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|}$ and $x^k_{t/k} := J^f_{t/k}\left(x^{k-1}_{t/k}\right)$, $x^0_{t/k} := x$, we have that:

$$d\left(S_t(x), \left(J^f_{t/k}\right)^k(x)\right) \le e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x^k_{t/k})} \qquad (2.10)$$

and similarly with $y^k_{t/k} := J^{f_n}_{t/k}\left(y^{k-1}_{t/k}\right)$, $y^0_{t/k} := y_n$, we also have that:

$$\begin{aligned}
d\left(\left(J^{f_n}_{t/k}\right)^k(y_n), S^n_t(y_n)\right) &\le e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f_n(y_n) - f_n(y^k_{t/k})}\\
&\le e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x^k_{t/k}) + \epsilon_n},
\end{aligned} \qquad (2.11)$$

where, in order to obtain the second inequality for fixed k, we used (M2) for $f_n(y_n) \to f(x)$, and (M1) for $f(x^k_{t/k}) \le \liminf_{n\to\infty} f_n(y^k_{t/k})$ where $y^k_{t/k} \to x^k_{t/k}$ as $n \to \infty$ by (2.9). Thus the remainder $\epsilon_n \to 0$.

Combining all of these estimates together with the a priori estimate of Lemma 1.4.2 for f, for a given $\epsilon > 0$, choose first a large enough k and then a large enough N to obtain that the right hand side of (2.7) is less than ϵ for all $n > N$. □

2.3 Sequences of spaces with an asymptotic relation

Let $(X_i, d_i), (X, d)$ be complete CAT(1)-spaces. Define:

$$\mathcal{X} := X \sqcup \left(\bigsqcup_i X_i \right)$$

as a set-theoretic disjoint union of topological spaces.

Definition 2.3.1 (Asymptotic relation, cf. [66], (Kuwae, 2008))**.** We call a topology on $\mathcal{X}$ an *asymptotic relation between* X_i and X if

(1) X_i and X are all closed in $\mathcal{X}$, and the restricted topology of $\mathcal{X}$ on each of X_i and X coincides with its original topology;

(2) for any $x \in X$ there exists a net $x_i \in X_i$ converging to x in $\mathcal{X}$;

(3) if $X_i \ni x_i \to x \in X$ and $X_i \ni y_i \to y \in X$ in $\mathcal{X}$, then we have $d_i(x_i, y_i) \to d(x, y)$;

(4) if $X_i \ni x_i \to x \in X$ and $y_i \in X_i$ is a net with $d_i(x_i, y_i) \to 0$, then $y_i \to x$ in $\mathcal{X}$.

In this section, we assume that X_i and X have an asymptotic relation. The following definition can be found in [66] (Kuwae, 2008), which generalizes weak convergence to an asymptotic relation in the CAT(0) setting. We modify it suitably for our CAT(1)-setting.

Definition 2.3.2 (Weak convergence)**.** A net $x_i \in X_i$ *weakly converges* to a point $x \in X$, that is $x_i \xrightarrow{w} x$, if for any net of π-convex geodesic segments $\gamma_i \subseteq X_i$ is strongly converging to a π-convex geodesic segment $\gamma \in X$ with $\gamma(0) = x$, such that, $d_i(\gamma_i, x_i) := \inf_{z \in \gamma_i} d_i(z, x_i) < \pi/2$, we have that $P_{\gamma_i}(x_i)$ it strongly converges to x. Strong convergence of γ_i to γ means that $\gamma_i(t) \to \gamma(t)$ for each $t \in D(\gamma)$.

Similarly to the case of a single CAT(1)-space, it is easy to prove that strong convergence implies weak convergence and that weak limit points are unique in the small enough metric balls.

Let $D \subset M^2_\kappa$ be a closed Jordan domain, whose boundary has finite length. Let X be a metric space. A map $f : D \mapsto X$ is *majorizing* if it is 1-Lipschitz and its restriction to the boundary ∂D is length-preserving. If Γ is a closed curve in X, we say that D majorizes Γ if there is a majorizing map $f : D \mapsto X$ such that the restriction $f|_{\partial D}$ traces out Γ. See Section 8.12 of [1] (Alexander, Kapovitch and Petrunin, 2019) for the following result.

Theorem 2.3.1 (Reshetnyak's majorization theorem)**.** For any closed curve Γ in a CAT(κ) space (of length at most $R_\kappa := \frac{\pi}{\sqrt{\kappa}}$ when $\kappa > 0$), there exists a convex region D in M^2_κ, and an associated map f such that D majorizes Γ under f.

The following three results appeared in [66] (Kuwae, 2008) as Lemma 5 (3-5) in the CAT(0)-setting and have also been mentioned in [65] (Kuwae, 2013) without proofs in the CAT(1)-case. We prove them below in the CAT(1)-setting.

Lemma 2.3.2. Let $x_i, y_i \in X_i$ and $x, y \in X$ such that $d_i(x_i, y_i), d(y, x) < \pi/2$ for all i. Assume $x_i \overset{w}{\to} x$ and $y_i \to y$. Then we have:

(1) $d(x, y) \leq \liminf_i d_i(x_i, y_i)$,

(2) $d_i(x_i, y_i) \to d(x, y)$ if and only if $x_i \to x$.

Proof. (1) : Take a net $\hat{x}_i \to x$ such that all geodesic segments $[\hat{x}_i, y_i]$ are π-convex. By the obtuse-angle property the spherical law of cosines gives

$$\begin{aligned}
&\cos d_i(x_i, y_i) \leq \cos d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \cos d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)) \\
&\quad + \sin d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \sin d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)) \cos \angle_{P_{[\hat{x}_i, y_i]}(x_i)}(x_i, y_i) \\
&\leq \cos d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \cos d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)) \\
&\leq \cos d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)),
\end{aligned}$$

which implies.

$$d_i(x_i, y_i) \geq d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)),$$

Thus,

$$\liminf_i d_i(x_i, y_i) \geq \liminf_i d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)) = d(y, x).$$

(2) : The implication "$\Leftarrow$" is obvious. We shall prove "$\Rightarrow$". The assumption yields that in the proof of (1) instead of inequalities we have equalities, thus it follows that $\cos d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \to 1$ implying $d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \to 0$. Since $P_{[\hat{x}_i, y_i]}(x_i) \to x$, this implies $x_i \to x$. □

Lemma 2.3.3. Given $\epsilon > 0$, let $x_i \in X_i$ be a net and $\gamma_i, \sigma_i : [0, 1] \to X_i$ geodesic segments such that x_i, γ_i, σ_i are contained in $(\pi/2 - \epsilon)$-convex sets for each index i and

$$\lim_i d_i(\gamma_i(0), \sigma_i(0)) = \lim_i d_i(\gamma_i(1), \sigma_i(1)) = 0. \tag{2.12}$$

Then:

$$\lim_i d_i(P_{\gamma_i}(x_i), P_{\sigma_i}(x_i)) = 0.$$

Proof. The $(\pi/2 - \epsilon)$-convexity assumption and (2.12) with Theorem 2.3.1 assures that

$$\lim_i \sup_{t \in [0,1]} d_i(\gamma_i(t), \sigma_i(t)) = 0. \tag{2.13}$$

Alternatively, in order to obtain (2.13) instead of Theorem 2.3.1 one can also use Ohta's L-convexity [87] (Ohta, 2007) available in CAT(1)-spaces. We set $y_i := P_{\gamma_i}(x_i)$, $z_i :=$

$P_{\sigma_i}(x_i)$ and take $s_i, t_i \in [0,1]$ satisfying $\gamma_i(s_i) = y_i$ and $\sigma_i(t_i) = z_i$. Then (2.13) leads to the equation:

$$\lim_i d_i(\sigma_i(s_i), y_i) = \lim_i d_i(\gamma_i(t_i), z_i) = 0.$$

In 2.6(1) [27] (Bridson and Haefliger, 1999) P is a non-expansive map. Thus $d_i(x_i, y_i) \leq d_i(x_i, \gamma_i(t_i))$ and $d_i(x_i, z_i) \leq d_i(x_i, \sigma_i(s_i))$. We have:

$$\lim_i |d_i(x_i, y_i) - d_i(x_i, z_i)| = 0,$$
$$\lim_i |d_i(x_i, y_i) - d_i(x_i, \gamma_i(t_i))| = 0.$$

By the spherical law of cosines we have:

$$\cos d_i(x_i, \gamma_i(t_i)) \leq \cos d_i(x_i, y_i) \cos d_i(y_i, \gamma_i(t_i))$$

which implies

$$\cos d_i(x_i, \gamma_i(t_i)) - \cos d_i(x_i, y_i) \leq \cos d_i(x_i, y_i)(\cos d_i(y_i, \gamma_i(t_i)) - 1),$$

Based on the trigonometric formula of cos, we have that $\cos d_i(x_i, \gamma_i(t_i)) - \cos d_i(x_i, y_i) \to 0$ since $d_i(x_i, y_i) - d_i(x_i, \gamma_i(t_i)) \to 0$. Furthermore, by the $(\pi/2 - \epsilon)$-convexity assumption we have:

$$\cos d_i(x_i, \gamma_i(t_i)), \cos d_i(x_i, y_i) \geq s$$

for some $s > 0$. Hence the above implies $0 \geq (\cos d_i(y_i, \gamma_i(t_i)) - 1) \to 0$, or equivalently, $d_i(y_i, \gamma_i(t_i)) \to 0$. Thus also $d_i(y_i, z_i) \to 0$. □

In the proof of the result below, we assume that the spaces X_i, X are separable. However, the argument should generalize to nets in non-separable spaces, since the main tool of Cantor's diagonalization process is available there as well. See for example [11] (Azouzi, 2023). The proof itself closely follows its CAT(0) variant of Lemma 5.5 in [66] (Kuwae, 2008).

Lemma 2.3.4. Let $\epsilon > 0$ be arbitrary. Any $x_i \in X_i$ net satisfying $d_i(x_i, o_i) < \pi/4 - \epsilon$ for a net $o_i \to o \in X$ has a weakly convergent subnet.

Proof. We may assume that $\{x_i\}$ is a countable sequence. First, take a dense countable subset $\{\xi_\nu\}_{\nu \in \mathbb{N}} \subseteq B_o(\pi/4 - \epsilon) \subseteq X$. Let $y_0 \in B_o(\pi/4 - \epsilon)$ and select $y_{0,i} \in B_{o_i}(\pi/4 - \epsilon) \subseteq X_i$ such that $y_{0,i} \to y_0$ strongly. For each $\nu \in \mathbb{N}$, we select a sequence $\xi_{\nu,i} \in B_{o_i}(\pi/4 - \epsilon) \subseteq X_i$ so that $\xi_{\nu,i} \to \xi_\nu$ strongly. Then, we have $\gamma^0_{\nu,i} := [y_{0,i}, \xi_{\nu,i}] \to [y_0, \xi_\nu] =: \gamma^0_\nu$ as $i \to \infty$. Since $d_i(x_i, y_{0,i})$ is bounded, $w^0_{\nu,i} := P_{[y_{0,i}, \xi_{\nu,i}]}(x_i)$ has a convergent subsequence whose limit denoted by w^0_ν is a point in $[y_0, \xi_\nu]$. By Cantor's diagonalization process, we can choose a common subsequence of $\{i\}$ independent of $\nu \in \mathbb{N}$ for which $w^0_{\nu,i} \to w^0_\nu$. We denote this subsequence with the indices $\{i\}$. Select an arbitrary sequence $0 < \epsilon_m \to 0$. We define $y_{m,i}$ and y_m inductively as follows. We assume that $y_{m,i}$ and y_m are defined so that $y_{m,i} \to y_m$. The sequence $w^m_{\nu,i} := P_{[y_{m,i},\xi_{\nu,i}]}(x_i)$ has a convergent subsequence, which can again be chosen independent of $\nu \in \mathbb{N}$ and we replace $\{i\}$ with that subsequence and set $w^m_\nu := \lim_i w^m_{\nu,i}$. There exists a number $\nu(m+1) \in \mathbb{N}$ such that:

$$d(y_m, w^m_{\nu(m+1)}) > \sup_\nu d(y_m, w^m_\nu) - \epsilon_m.$$

where, $y_{m+1} := w^m_{\nu(m+1)}$ and $y_{m+1,i} := w^m_{\nu(m+1),i}$. Then, $y_{m,i} \to y_m$ for each m.

By the obtuse-angle property, the spherical law of cosines implies:

$$\cos d_i(y_{m,i}, x_i) \leq \cos d_i(x_i, y_{m+1,i}) \cos d_i(y_{m+1,i}, y_{m,i}).$$

Taking a subsequence of $\{i\}$ again, we can assume that for each m, $d_i(y_{m,i}, x_i)$ converges to some $\lambda_m \in \mathbb{R}$. This implies that λ_m is monotone non-increasing, hence $d(y_{m+1}, y_m) \to 0$ as well. For any $\nu \in \mathbb{N}$,

$$\lim_i d_i(y_{m,i}, w^m_{\nu,i}) = d(y_m, w^m_\nu) < \epsilon_m + d(y_{m+1}, y_m) =: \epsilon'_m \to 0. \tag{2.14}$$

There exists $\{\nu(l,i)\}$, such that $\lim_i \xi_{\nu(l,i)} = y_l$. By (2.14), if $\nu \ll i$, then $d_i(y_{m,i}, w^m_{\nu,i}) \leq \epsilon'_m$. We may assume $\nu(l,i) \ll i$, so that we have:

$$\limsup_i d_i(y_{m,i}, w^m_{\nu,i}) \leq \epsilon'_m.$$

Since $[y_{m,i}, \xi_{\nu(l,i),i}] \to [y_m, y_l]$, we can choose a common subsequence $\{i\}$ independent of m, l for which $w^m_{\nu(l,i),i} \in [y_{m,i}, \xi_{\nu(l,i),i}]$ converges to some point $x_{m,l} \in [y_m, y_l]$. By the inequality above:

$$d(y_m, x_{m,l}) \leq \epsilon'_m.$$

Applying Lemma 2.3.3 to $\gamma_i := [y_{m,i}, y_{l,i}]$ and $\sigma_i := [y_{m,i}, y_{v(l,i),i}]$, we have $P_{[y_{m,i},y_{l,i}]}(x_i) \to x_{m,l}$. Therefore,

$$d(y_l, y_m) \leq d(y_l, x_{l,m}) + d(x_{l,m}, y_m) \leq \epsilon'_m + \epsilon'_l$$

implying that $\{y_m\}$ is Cauchy. Let $x := \lim_m y_m$. By (2.14), for any geodesic γ emanating from x with an associated sequence of geodesics $\gamma_i \in X_i$ strongly converging to γ we have that $P_{\gamma_i}(x_i) \to x$ strongly. Then it follows that $x_i \xrightarrow{w} x$ completing the proof. □

Remark 2.3.1. It is interesting to compare the statement of the above Lemma 2.3.4 with Lemma 2.1.2. Can we have the weak compactness of larger sets as well in the first mentioned 2.3.4 Lemma such as in Lemma 2.1.2 for a single space? Note that the proof above uses projections in an essential way, on the other hand, the asymptotic center approach in [106] (Yokota, 2016) can prove Lemma 2.1.2 in a different way obtaining weak compactness in π-convex sets.

Definition 2.3.3 (Mosco convergence for asymptotic relations)**.** A sequence of lsc functions $f_n : X_n \mapsto \overline{\mathbb{R}}$ said to converge to $f : X \mapsto \overline{\mathbb{R}}$ in the sense of Mosco if, for any $x \in X$, we have

(M1) $f(x) \leq \liminf_{n\to\infty} f_n(x_n)$ whenever $x_n \in X_n$ and $x_n \xrightarrow{w} x$,

(M2) there exists a sequence $y_n \in X_n$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$.

We prove a version of Proposition 2.2.7 adapted to the setting of this section.

Proposition 2.3.5 (Ekeland principle, bounded case)**.** Let $\operatorname{diam}(X) < \pi/2 - \epsilon$ for an $\epsilon > 0$. Given $w_n \in X_n$ such that $w_n \to w \in X$ and a sequence of lsc λ-convex functions $f_n : X_n \mapsto (-\infty, \infty]$ that is Mosco converging to $f : X \mapsto (-\infty, \infty]$. Then there exist $\alpha, \beta \geq 0$ such that

$$f_n(x_n) \geq -\alpha d_n(x_n, w_n) - \beta$$

for all $x_n \in X_n$ and $n \in \mathbb{N}$.

Proof. Assume that the assertion is false; that is, for any $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ and $x_{n_k} \in X_{n_k}$ such that

$$f_{n_k}(x_{n_k}) + k[d_{n_k}(x_{n_k}, w_{n_k}) + 1] < 0.$$

We first assume that $n_k \to \infty$ as $k \to \infty$. Since we have $\operatorname{diam}(X) < \pi$, we can select a weakly convergent subsequence from x_{n_k} still denoted by x_{n_k} with weak limit denoted by $\overline{x} \in X$. Then by the Mosco convergence of f_n we have:

$$\begin{aligned} f(\overline{x}) \leq \liminf_{k\to\infty} f_{n_k}(x_{n_k}) &\leq \liminf_{k\to\infty} -k[d_{n_k}(x_{n_k}, w_{n_k}) + 1] \\ &\leq -\limsup_{k\to\infty} k[d_{n_k}(x_{n_k}, w_{n_k}) + 1] \\ &= -\infty \end{aligned}$$

which is a contradiction. The rest of the proof is similar to the proof of Proposition 3.7. □

Theorem 2.3.6. Let $\operatorname{diam}(X) < \pi/2 - \epsilon$ for an $\epsilon > 0$ and $f_n : X_n \to (-\infty, \infty]$ a sequence of lsc λ-convex functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Then,

$$\lim_{n\to\infty} (f_n)_\tau(x_n) = f_\tau(x)$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x_n) = J_\tau^f(x)$$

for any small enough $\tau > 0$ and $x \in D(f)$ and $x_n \in X_n$ with $x_n \to x$.

Proof. From Proposition 2.3.5 we have:

$$f_n(J_\tau^{f_n}(x_n)) \geq -\alpha d_n(J_\tau^{f_n}(x_n), w_n) - \beta.$$

From the definition of $J_\tau^{f_n}(x_n)$ and (M2), we have:

$$f_n(y_n) + \frac{1}{2\tau} d_n^2(x_n, y_n) \geq f_n(J_\tau^{f_n}(x_n)) + \frac{1}{2\tau} d_n^2(x_n, J_\tau^{f_n}(x_n))$$

where $y_n \to x$, which combined with the above estimate yields:

$$f_n(y_n) + \frac{1}{2\tau} d_n^2(x_n, y_n) + \alpha d_n(J_\tau^{f_n}(x_n), w_n) + \beta \geq \frac{1}{2\tau} d_n^2(x_n, J_\tau^{f_n}(x_n)).$$

That is,

$$0 \geq d_n^2(x_n, J_\tau^{f_n}(x_n)) - 2\tau\alpha d(J_\tau^{f_n}(x_n), w_n) - 2\tau(f_n(y_n) + \beta) - d_n^2(x_n, y_n).$$

Since $d_n^2(x_n, y_n) \to 0$ and $f_n(y_n) \to f(x) < \infty$ by (M2), for small enough $\tau > 0$ the above forces

$$d_n(x_n, J_\tau^{f_n}(x_n)) < s \tag{2.15}$$

for all large enough $n \in \mathbb{N}$, given an arbitrary $s \leq \frac{\pi}{4} - \epsilon$. Now, we are in position to prove $\lim_{n\to\infty} J_\tau^{f_n}(x_n) = J_\tau^f(x)$. Pick any subsequence $j_k \in J_\tau^{f_{n_k}}(x_{n_k})$. Since $\operatorname{diam}(X) < \pi/2 - \epsilon$, there exists a subsequence of $\{j_k\}_{k\in\mathbb{N}}$ still denoted by j_k with weak limit $c \in X$. Since $f_n \to f$ in the sense of Mosco, there exists a sequence $y_n \to J_\tau^f(x)$ and $f_n(y_n) \to f(J_\tau^f(x))$. Then using (1) of Lemma 2.3.2 we get:

$$\begin{aligned}
\limsup_{k\to\infty} (f_{n_k})_\tau(x_{n_k}) &\leq \limsup_{k\to\infty} f_{n_k}(y_{n_k}) + \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, y_{n_k}) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau} d^2(x, J_\tau^f(x)) \\
&\leq f(c) + \frac{1}{2\tau} d^2(x, c) \\
&\leq \liminf_{k\to\infty} f_{n_k}(j_k) + \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, j_k) \\
&= \liminf_{k\to\infty} f_{n_k}(J_\tau^{f_{n_k}}(x_{n_k})) + \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, J_\tau^{f_{n_k}}(x_{n_k})) \\
&= \liminf_{k\to\infty} (f_{n_k})_\tau(x_{n_k})
\end{aligned} \tag{2.16}$$

which yields $c \in J_\tau^f(x)$. Furthermore using (M1) and (M2) with $z_n \to c$ and $f_n(z_n) \to$

$f(c)$ we have:

$$\begin{aligned}
\limsup_{k\to\infty} \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, j_k) &\leq \limsup_{k\to\infty} -f_{n_k}(j_k) + f_{n_k}(z_{n_k}) + \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, z_{n_k}) \\
&= -\liminf_{k\to\infty} f_{n_k}(j_k) + f(c) + \frac{1}{2\tau} d^2(x, c) \\
&\leq -f(c) + f(c) + \frac{1}{2\tau} d^2(x, c) \\
&\leq \frac{1}{2\tau} d^2(x, c) \\
&\leq \liminf_{k\to\infty} \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, j_k)
\end{aligned}$$

which together with (2.15) and (2) of Lemma 2.3.2 prove the strong convergence:

$$j_k \to c.$$

Then using (2.16) we get:

$$\begin{aligned}
\lim_{n\to\infty} (f_n)_\tau(x_n) &= \lim_{n\to\infty} f_n(J_\tau^{f_n}(x_n)) + \frac{1}{2\tau} d_n^2(x_n, J_\tau^{f_n}(x_n)) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau} d^2(x, J_\tau^f(x)) = f_\tau(x)
\end{aligned}$$

completing the proof of the second part of the assertion. □

At this point the analogues of Theorem 2.2.9 or Theorem 2.2.10 for asymptotic relations can be established as well in the same manner. We omit the details.

Theorem 2.3.7. Assume $f_n : X_n \to (-\infty, \infty]$ is a sequence of lsc semi-convex functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Let $J_\tau^f(x)$ and $J_\tau^{f_n}(x_n)$ be the corresponding resolvents such that for all small enough $\tau > 0$:

$$J_\tau^{f_n}(x_n) \to J_\tau^f(x)$$

as established earlier in this section and let $S_t(x)$ and $S_t^n(x_n)$ denote the corresponding gradient flows. Then:

$$\lim_{n\to\infty} S_t^n(x_n) = S_t(x)$$

for any $t \geq 0$ and $x \in D(f)$ and $x_n \in X_n$ with $x_n \to x$.

Proof. By (M2) there exists an $y_n \to x$ with $f_n(y_n) \to f(x)$. Since $y_n, x_n \to x$, thus $d_n(y_n, x_n) \to 0$ implying

$$d_n(S_t^n(y_n), S_t^n(x_n)) \leq e^{-\lambda t} d_n(y_n, x_n) \to 0,$$

so it is enough to show that $S_t^n(y_n) \to S_t(x)$. First, we claim that:

$$\left(J_{t/k}^{f_n}\right)^k (y_n) \to \left(J_{t/k}^{f}\right)^k (x) \tag{2.17}$$

for any fixed $k \in \mathbb{N}$. Indeed, $y_n \to x$ and assuming that (2.17) holds for a $k \geq 0$, Theorem 2.3.6 proves (2.17) for $k+1$ yielding the claim by induction.

Next, by (1.6.2) with $C_0 := \sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|}$ and $x_{t/k}^k := J_{t/k}^f\left(x_{t/k}^{k-1}\right)$, $x_{t/k}^0 := x$ we have that:

$$d\left(S_t(x), \left(J_{t/k}^{f}\right)^k (x)\right) \leq e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x_{t/k}^k)}. \tag{2.18}$$

Similarly with $y_{t/k}^k := J_{t/k}^{f_n}\left(y_{t/k}^{k-1}\right)$, $y_{t/k}^0 := y_n$ we also have that:

$$\begin{aligned} d_n\left(\left(J_{t/k}^{f_n}\right)^k (y_n), S_t^n(y_n)\right) &\leq e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f_n(y_n) - f_n(y_{t/k}^k)} \\ &\leq e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x_{t/k}^k) + \epsilon_n}, \end{aligned} \tag{2.19}$$

where we used (M2) for $f_n(y_n) \to f(x)$, and (M1) for $f(x_{t/k}^k) \leq \liminf_{n\to\infty} f_n(y_{t/k}^k)$ where $y_{t/k}^k \to x_{t/k}^k$ as $n \to \infty$ by (2.17) to obtain the second inequality for fixed k. Thus the remainder $\epsilon_n \to 0$.

Now, choose sequences $z_n^t, j_n^{t,k} \in X_n$ such that $z_n^t \to S_t(x)$ and $j_n^{t,k} \to \left(J_{t/k}^{f}\right)^k (x)$ as $n \to \infty$. Then, it is enough to show that $d_n(S_t^n(y_n), z_n^t) \to 0$. We estimate as:

$$\begin{aligned} d_n(S_t^n(y_n), z_n^t) \leq\; & d_n\left(S_t^n(y_n), \left(J_{t/k}^{f_n}\right)^k (y_n)\right) \\ & + d_n\left(\left(J_{t/k}^{f_n}\right)^k (y_n), j_n^{t,k}\right) + d_n\left(j_n^{t,k}, z_n^t\right) \end{aligned} \tag{2.20}$$

Combining all these estimates together with the a priori Lemma **??** for f and an arbitrary $\epsilon > 0$, we first choose a large enough k, so that the right hand side of (2.18) is less than ϵ, which implies that there exists a large enough $N_1 > 0$, such that the last term on the right hand side of (2.20) is less than 3ϵ for all $n > N_1$. Then, choose a large enough $N_2 > N_1$ such that for all $n > N_2$ by (2.17) the second term on the right hand side of (2.20) is less than 2ϵ. Lastly, choose a large enough $N_3 > N_2$, such that for all $n > N_3$ the right hand side of (2.19) is less than 2ϵ making the first term on the right hand side of (2.20) less than 2ϵ. This implies that $d_n(S_t^n(y_n), z_n^t) < 7\epsilon$, thus the desired $S_t^n(y_n) \to S_t(x)$ follows. □

2.4 Conclusion on Mosco convergence of semi-convex functions in CAT(1)-space

We use weak compactness to find a weakly convergent subsequence of $J_\tau^{f_n}(x)$ indexed by n. Then we show that the limit of this subsequence is actually $J_\tau^f(x)$, and then that the whole sequence actually strongly converges to $J_\tau^f(x)$. Among others, we also use an Ekeland type of lower estimate on $f_n(x)$ to do this, that we had to prove as well.

We also proved the same result for Mosco converging sequences of lsc semi-convex L-Lipschitz functions and Mosco converging sequences of lsc semi-convex functions that are uniformly bounded below by 0. The definition of weak convergence and compactness considered here agrees with the usual weak convergence and weak sequential compactness of bounded sets in Hilbert spaces.

For illustrative purposes and better understanding, imagine a sequence of valleys converging to a valley, and on them steepest descent curves. The valleys symbolize the functions f_n, while the steepest descent curves symbolize the gradient flows. We expect that if the valleys deform continuously then the corresponding steepest descent curves deform continuously as well. This is the intuition behind the idea that Mosco convergence should imply the convergence of corresponding gradient flows.

We conclude that the commutativity property and the semi-convexity of the squared distance function are enough to establish the Evolution Variational Inequalities (EVI) and the contractivity of the gradient flow. Mosco convergence implies convergence of resolvents, which in turn implies convergence of gradient flows on the CAT(1)-setting. A fundamental difference between the CAT(0) and CAT(1)-space is that the CAT(1)-space lacks 1-convexity, thus all analysis is local under weaker semi-convexity properties, while for the CAT(0)-space all analysis can be carried out globally due to the global 1-convexity of the squared distance function. There are a number of other significant mathematical alterations in the CAT(1)-space, which render it differently from the CAT(0)-setting.

CHAPTER 3

Convergence of Douglas-Rachford Operator Splitting Algorithm

The current chapter aims to examine the convergence of the Douglas Rachford operator splitting algorithm, which is an *iteration process* of minimization for two convex functions. Such as the aim is the minimization of transaction costs and we assume that the investor makes a financial decision in the second period. The investor during the first period collects market information and simultaneously develops financial skills. From mathematical and technical point of view, the goal is to find a zero of the sum of two maximal monotone operators according to [75] Luenberger (1969), [93] Parikh and Boyd (2013).

The research discusses first a theoretical model and then a numerical analysis from finance. In detail, the Douglas-Rachford convergence of Moreau envelopes through the lens of portfolio optimization model for a selected financial assets. The investor through self-controlling and self-development behavior creates the best possible financial strategy. This cautious behavior leads the agent to invest in the second period, not in the first. Lions and Mercier [72] (Lions and Mercier, 1979) have verified the convergence of non-linear maximal monotone dual operators, while Svaiter [100] (Svaiter, 2019) has been proved the weak convergence of the Douglas Rachford method [100] (Svaiter, 2019). In the current work, the aspects of qualitative and quantitative analysis provide a comparison estimation of the standard Douglas Rachford algorithm.

This part relates to the previous chapter regarding convergence, where in the first chapter, the sequence of semi-convex and lower semi-continuous functions are weakly converging in the sense of Mosco. At the same time, the current segment is associated with the strong convergence estimation of two convex and continuous functions. The common application, which was used in both chapters is the *Moreau-Yosida envelop* *theorem* described later in the Definition of 3.4.1.

A number of *pre-conditions* of the Douglas Rachford method were listed by Stephen Boyd and Pontus Giselsson [48] (Boyd and Giselsson, 2014) in their work of *Diagonal Scaling in Douglas-Rachford Splitting and ADMM*, where the abbreviation of ADMM stands for the Alternating Direction Method of Multipliers. Another group of authors dealing with the convergence rate, such as Lions and Mercier [72] (Lions and Mercier, 1979) have verified the convergence of non-linear maximal monotone dual operators, also Svaiter has proved the weak convergence of the Douglas Rachford method [100] (Svaiter, 2019). In the more recent research, considering the last five years, we can observe some specific versions of Douglas-Rachford operation splitting, such as Convergence Rates of *Forward–Douglas–Rachford Splitting Method* by Molinari, Liang and Fadili [79] (Fadili, Molinari and Liang, 2019), similarly Convergence of an *Inertial Shadow Douglas-Rachford Splitting Algorithm for Monotone Inclusions* by Fan, Qin and Tan [40] (Fan, Qin and Tan, 2021).

3.0 Motivation

Optimal minimization of transaction costs, in the form of a portfolio bid-ask spread can be obtained by the analytical and numerical investigation. The **research question** relates to the inquiry: What is the convergence behaviour in the case of Douglas-Rachford iteration? The **methodology** involves numerical experiments on a financial application, incorporating the portfolio risk factor in the estimation.

3.1 Preliminaries of Dual Optimization

The concept of *duality principle* for optimization involves conjugate convex or concave functionals. Boyd and Ryu wrote a survey on monotone operator methods in optimization [97] (Ryu and Boyd, 2016). The assumption relies on searching for a standard minimization in association with two convex functionals over a convex domain.

Definition 3.1.1 (Conjugate Convex Functionals)**.** Let f be a convex functional defined on a convex set C in a normed space, the conjugate set C^* is defined as

$$C^* = \left\{ x^* \in X^* : \sup_{x \in C} [\langle x, x^* \rangle - f(x)] < \infty \right\}$$

It can be seen that the structure of dual optimization is easy to invert $(I + \lambda D + \lambda F + \lambda^2 DF)$, because the latter is simply $(I + \lambda D)(I + \lambda F)$. This motivates further the

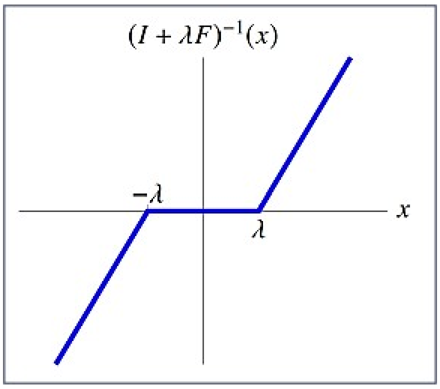

Figure 3.1: Fixed time step of λ parameter

backward difference formula according to the Backward Euler:

$$x^{n+1} = x^n + \lambda(-Dx^{n+1} - Fx^n)$$

According to the backward difference formula in the context of Douglas-Rachford [35] (Douglas and Rachford, 1956) it proceeds:

$$x^{n+1} = x^n + \lambda(-Dx^{n+1} - Fx^{n+1}) + \lambda^2 DF(x^n - x^{n+1})$$

This can be rewritten as $(I + \lambda D)(I + \lambda F)x^{n+1} = x^n + \lambda^2 DFx^n$. See Figure 3.1. of fixed time step of λ parameter $(I + \lambda F)^{-1}$.

The latter applies to the Laplacian and any positive linear operators as D and F. Lions and Mercier [72] (Lions and Mercier, 1979) proved that the previous application is convergent to non-linear maximal monotone operators, such as Douglas-Rachford in our example can be rewritten as follows:

$$x^{n+1} = Ex^n$$

where, $E = (I+\lambda D)^{-1}(I+\lambda F)^{-1}(I+\lambda^2 DF)$. Further, $I+\lambda^2 DF = (I+\lambda D)\lambda F+I-\lambda F$, it results:

$$E = (I + \lambda F)^{-1}[I + \lambda D)^{-1}(I + \lambda F) + \lambda F]$$

If $(I + \lambda F)^{-1}x^{*,n} = x^n$, then we can get:

$$x^{*,n+1} = ((I + \lambda D)^{-1}(I - \lambda F) + \lambda F(I + \lambda F)^{-1}x^{*,n}$$
$$= [(I + \lambda D)^{-1}(2(I + \lambda F)^{-1} - I) + I - (I + \lambda F)^{-1}]x^{*,n}$$

where, $(I - \lambda F)(I + \lambda F)^{-1} = 2(I + \lambda F)^{-1} - I$ and $\lambda F(I + \lambda F)^{-1} = I - (I + \lambda F)^{-1}$. In notation, $(I + \lambda D)^{-1} = prox_{\lambda f}$ and $(I + \lambda F)^{-1} = prox_{\lambda g}$, which are the proximal operators associated with the lsc functions. In order to minimize the latter two convex

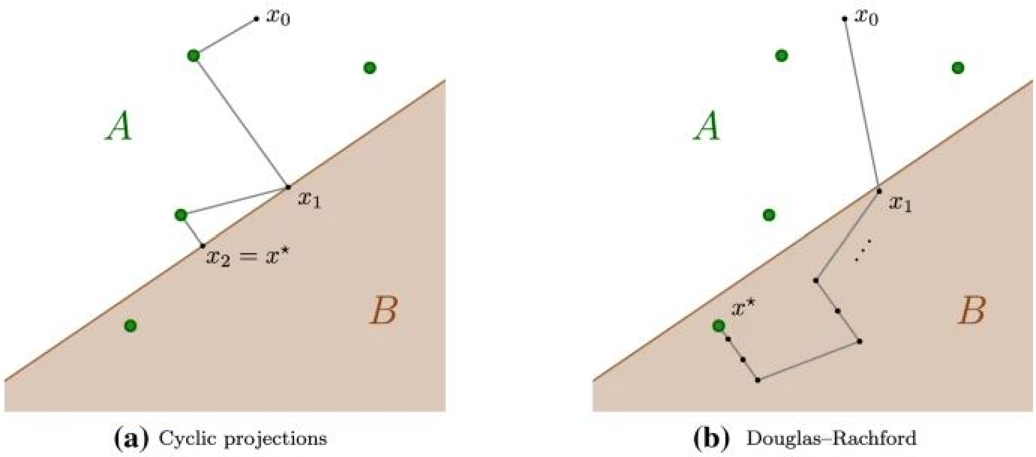

Figure 3.2: Cyclic projection vs. Douglas Rachford. *Source: Aragón et al.(2020)*

functions $f(x)$ and $g(x)$, it goes through the following process of iteration [7]:

$$\begin{aligned} x^{n+1} &= \operatorname{prox}_f(y_n - z_n) \\ y^{n+1} &= \operatorname{prox}_g(x^{n+1} + z_k) \\ z^{n+1} &= z_k + (x^{n+1} - y_{n+1}) \end{aligned}$$

Figure 3.3. indicates the Douglas–Rachford method for solving the best approximation, where the set of A and B are closed subsets (left side) and the algorithm itself converges from x_0 to x^*. In comparison, we can see a halfspace (right side), where the convergence iterates toward x^*.

Without loss of generality, the Douglas-Rachford operator is a sum of two convex functions as $x^{n+1} = Ex^n$, it results:

$$E(x) = \operatorname{prox}_f(2\operatorname{prox}_g(x) - x) + x - \operatorname{prox}_g(x)$$

Definition 3.1.2. A map E: $R^n \to R^n$ is *firmly non-expansive* if:

$$\| E(x) - E'(x) \|_2^2 \leq \langle x - x^*, E(x) - E(x^*) \rangle, \quad \forall x, x^* \in R^n.$$

The proximal operators of convex functions are firmly non-expansive if the proximal operator is closed, convex and proper. The sequence of iteration originally generated by the proximal operator $x^n + 1 = prox_f(x_n))$ converges to a fixed point, $x = prox_f(x)$, which is the minimizer of f. The subgradient property ∂f here indicates the relationship between the point x and the proximal operator of a function.

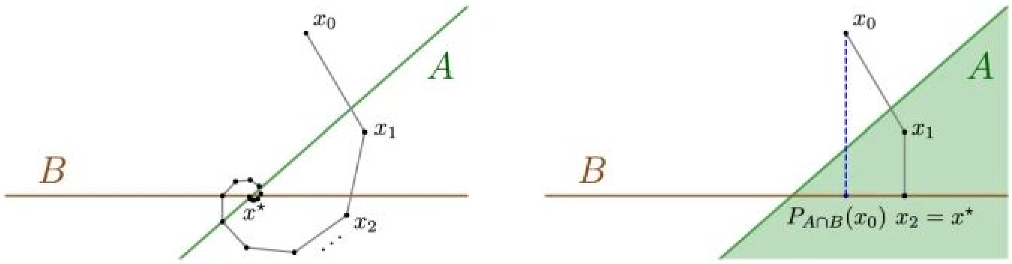

Figure 3.3: Douglas–Rachford method for solving the best approximation. *Source: Aragón et al.(2020)*

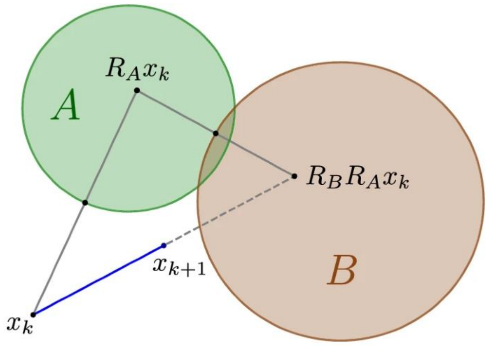

Figure 3.4: Geometric interpretation of the Douglas–Rachford iteration. *Source: Aragón et al.(2020)*

3.2 Preconditions of Douglas-Rachford operator splitting

This part examines those *preconditions*, which are necessary to perform an iterative method of optimization. We rely on a work of [48] Giselsson and Boyd (2014), who improved the generalized Douglas-Rachford algorithm under the following *precursory conditions*:

(a) *Smoothness and strong convexity:*

$$\begin{aligned} f(x) &\geq f(y) + \langle u, x-y \rangle + \frac{1}{2}\|x-y\|^2 \\ &= f(Pr) + \langle P^{-1}w, P(q-r) \rangle + \frac{1}{2}\|Pr - Pq\|^2 \\ &= f_P(r) + \langle w, q-r \rangle + \frac{1}{2}\|r-q\|^2 \end{aligned}$$

where P is a *diagonal preconditioner*, which has a role to use a technique that approximates the inverse of a matrix. The diagonal elements of matrix further accelerates the convergence of this iterative method). Besides, [26] Kristian Bredies and Hongpeng Sun (2015) applied *various efficient, linear and continuous preconditioners* for proposing a version of the Douglas-Rachford splitting method for solving convex-concave saddle-point problems associated with Fenchel-Rockafellar duality [26] (Bredies and Sun, 2015). Similarly, Enis Chenchene, Kristian Bredies and Emanuele Naldi (2022) [25] proposed a few graph-based consequences of the Douglas-Rachford splitting (DRS) in order to solve monotone inclusion problems and it can be understood as a pattern of a Preconditioned Proximal Point (PPP) method, which ensures robust convergence.

Technically, the DRS inquires a solution at each step-size of iteration with the aim to estimate the resolvent operators.

3.3 Lie-Trotter-Kato formula in the operator topology

The below convergence of infinite dimension is related to the *strong operator topology* on a Hilbert space, where the locally convex topology on the set of bounded operators are induced by the semi-norms. The Lie-Trotter-Kato product formula for linear operators $F \circ D = D \circ F$, in the current case can be rewritten:

$$\lim_{n\to\infty} \left(e^{\frac{F}{n}} \cdot e^{\frac{D}{n}} \right)^n = e^{F+D} \tag{3.1}$$

which is the analogue form of a *classical exponential law of complex numbers* $e^F \cdot e^D = e^{F+D}$. Onwards, let's see the separate contribution of each author: Lie, Trotter and Kato. According to Zagrebnov, the Trotter product formula on Hilbert and Banach spaces for operator-norm convergence we can [107] (Zagrebnov, 2020) rewrite on the current case:

- **S.Lie (1875):** The *product formula* for matrices F and D:

$$e^{-\tau H} = \lim_{n\to\infty} \left(e^{-\tau \frac{F}{n}} \cdot e^{-\tau \frac{D}{n}} \right)^n, \tag{3.2}$$

where, $\tau \geq 0$ and $F+D = H$ are the summary forms of operators F and D. The formula (3.3) can be implemented on a bounded linear operator $L(H)$ on a Hilbert space (H), and this operator is continuous. Since, we have the Lie formula, further the operator norm convergence rate can be easily obtained by reordering:

$$\sup_{\tau\in[0,T]} \left\| e^{-\tau \frac{F}{n}} \cdot e^{-\tau \frac{D}{n}} - e^{-\tau H} \right\|_{L(H)} = O(1/n) \tag{3.3}$$

- **H.F. Trotter (1959):** he has exposed unbounded operators on Banach spaces [102] (Trotter, 1959):

$$e^{-\tau H} = \mathrm{s} - \lim_{n\to\infty} \left(e^{-\tau \frac{F}{n}} \cdot e^{-\tau \frac{D}{n}} \right)^n \tag{3.4}$$

In the 1990s it was proved that the Lie-Trotter formula is not only valid in the case of strong operator topology related to the convergence of product formula, but also in the case of operator-norm convergence on a Hilbert space [85] (Zagrebnov and Niedhardt, 1999).

- **T. Kato (1978):** The below equation has been introduced by [60] [61] Kato (1974, 1978) for a positive self-adjoint operators in the context of arbitrary pair of self-adjoint contraction semigroups, where by reformulating on the current case, such as on the operator of F and D, the Trotter formula holds in the *strong operator topology* if $\mathrm{dom}(\sqrt{F}) \cap \mathrm{dom}(\sqrt{D})$ has been applied on Hilbert space [107] (Zagrebnov, 2020).:

$$e^{-\tau H} = \mathrm{s} - \lim_{n \to \infty} \left(e^{-\tau \frac{F}{n}} \cdot e^{-\tau \frac{D}{n}} \right)^n \tag{3.5}$$

Rogava [96] (Rogava, 1993) stated that if F and D are *positive self-adjoint operators*, such that $\mathrm{dom}(F) \subseteq \mathrm{dom}(D)$ and the summary form of the operators: $F+D = H$ is self-adjoint, then the following holds [84] (Zagrebnov and Niedhardt, 1998), [54] (Ichinose et al., 2001):

$$\left\| (e^{-\tau \frac{F}{n}} \cdot e^{-\tau \frac{D}{n}})^n - e^{-\tau H} \right\|_{L(H)} = O\left(\log(n)/\sqrt{n}\right) \tag{3.6}$$

as $n \to \infty$. Neidhardt and Zagrebnov [85] (Zagrebnov and Niedhardt, 1999) have introduced those *necessary and sufficient conditions*, which guarantee the convergence of operator-norm related to the Trotter-Kato product formula for given positive self-adjoint operators and the Kato-functions $f(\cdot)$ and $g(\cdot)$. Further, the convergence of the operator-norm in the context of the Trotter-Kato product formula holds if the resolvent of one of the operators is *compact*; or either operator is relatively compact with respect to another one; or if the product (the sum) of the resolvents of the involved operators is compact. By the consequence of Baker-Campbell-Hausdorff formula, the following equality holds: $e^F \cdot e^D = e^H$. The pair of positive self-adjoint operators is in the equation (3.6.), what we know that these operators are bounded, the resolvent is compact or relatively compact and, the metric space is infinite dimensional.

3.4 Douglas-Rachford Monotone Operators

Boyd, Donoghue, and Wang[24] (Boyd et al., 2012) discussed a multi-period portfolio optimization using ADMM. The second author Donoghue [86] (O'Donoghue et al., 2012) described the mentioned general splitting method with a high-frequency series in his work, by solving efficient quadratic control problems as optimization problems both in single and multi-period portfolio optimization models. The authors used a fixed-point arithmetic application on a Field Programmable Gate Array (FPGA). The latter method uses logical data, real-time market, and historical data. The algorithm provides evidence for future strategies in buy and sell actions, asset allocations, and risk management. FPGA also mitigates financial risk and time between the optimal decision and trade execution. The current research involves financial trading days in the numerical analysis, and based on real-time market data, it has an aim to minimize the transaction costs.

Through the minimization of two convex functions, f(x) and g(x), we can get the Douglas-Rachford method and let the $(x_t)_i$ denote the sequence of iterations of a specific portfolio.

Definition 3.4.1 (Douglas-Rachford operator)**.** The operator of convex functions $f(x)$ and $g(x)$ is firmly non-expansive, such as related to the proper, convex, and lsc functions. The operator D: $R^n \to R^n$ is a firmly non-expansive map, which possesses at least one fixed point $x*$ and by the iteration $x_{t+1} = D(x_t)$ it converges to some fixed point of D. Let's take the monotone operator:

$$D(x_t) = \mathrm{prox}_f(2 \cdot \mathrm{prox}_g(x_t) - x_t) + x_t - \mathrm{prox}_g(x_t) \tag{3.7}$$

If the D operator is firmly non-expansive, such as related to the proper, convex, and lsc functions, then $F = I - D$ is also firmly non-expansive as well. The proximal iteration enables that the convergence moves towards a fix point.

Let H be a real Hilbert space, and the evolution equation:

$$\frac{\partial f}{\partial t} + D(f) \in 0 \tag{3.8}$$

In the equation (3.8) D is a monotone operator and $f(0) = f^0$ and $f \in H$ [72] (Lions and Mercier, 1979). It is possible to reformulate the setup on operator F, where the evolution equation is: $\frac{\partial g}{\partial t} + F(g) \in 0$, if $g \in H$, where $g(0) = f^0$. The Douglas-Rachford operator splitting algorithm is defined with a fixed time step of λ, where the time step (λ) is sufficiently small if F and D operators are Lipschitz-continuous [49] (Goldstein, 1964). Then, the operators, F and D, can be rewritten according to a priori estimation [72] (Lions and Mercier, 1979) in the following structure for the financial application:

$$f^{n+1} = (I + \lambda F)^{-1}[(I + \lambda D)^{-1}(I - \lambda D)\lambda F]f^n \tag{3.9}$$

Further, the Douglas-Rachford operator of functions f and g can be defined as iterating the proximal mappings of P_f and P_g:

$$DR(f, g) := I - P_f + P_g \cdot R_f \tag{3.10}$$

where, $R_f := 2P_f - I$ - is a reflectant or reflected proximal mapping; important property of Douglas-Rachford (DR) is that, it is firmly non-expansive: $P_f + P_f* = I$. [17]

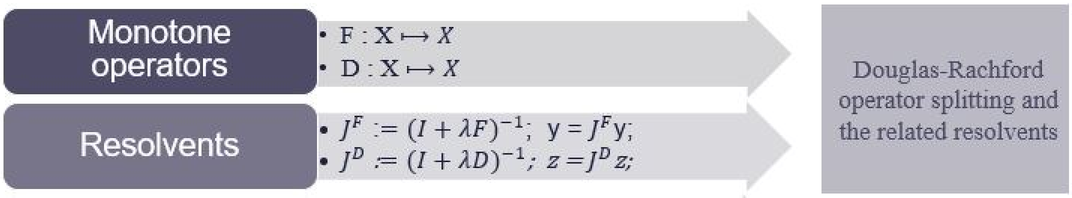

Figure 3.5: Composition of firmly non-expansive monotone operators and their resolvents

(Bauschke and Moursi, 2023), such as the subset is Lipschitz continuous with constant $L = 1$ or simply L-Lipschitz. Therefore, we say that the DR operator is non-expansive. If it would be $L < 1$, than it would be a Banach contraction property. It is very interesting in the work of Comebettes [30] (Combettes, 1964), that he particularly emphasizes the fact that the subdifferential of the proximal mapping of a convex function is actually the resolvent configuration of its subdifferential (Figure 2.2.).

3.5 Empirical finance

In empirical finance application, an investor tends to make a series of optimal decisions. This agent through self-controlling and self-development behavior creates the best possible financial strategy. This cautious behavior leads the agent to invest a posteriori and in the second period, not in the first. This paper reflects on the Douglas-Rachford Splitting Algorithm setting, which was developed by Jim Douglas and Henry Rachford [35] (Douglas and Rachford, 1956) in 1950s and represents a special form of Alternating Direction Method of Multipliers (ADMM). The latter method solves a convex optimization problem by performing a process of splitting into it's smaller parts. Through this splitting procedure of space variables, and according to various splitting schemes, it results different algorithms. The Douglas-Rachford Operator Splitting Algorithm was originally invented as a technique to solve a heat equation for space variables implementing finite difference scheme method. Later, Pierre Louis Lions and Bertrand Mercier [72] (Lions and Mercier, 1979) proved the convergence of non-linear maximal monotone dual operators in relation with the Douglas-Rachford splitting method.

In general, the risk means two aspects:

1. systematic risk;

2. non-systematic risk.

In the current case, we are dealing with the non-systematic risk on a portfolio level of investigation. In financial mathematics, the Douglas-Rachford operator splitting is one of the several splitting scheme possibilities:

- *Forward-Backward splitting* - represents two positive operators. A Forward-Euler as a first positive operator, and a Backward-Euler as a second positive operator in the process of operator splitting;

- *Peaceman-Rachford splitting* - is a pioneer form of Douglas-Rachford operator splitting, which requires solving a linear system [72] (Lions and Mercier, 1979) at each time step for both reflectant operator [17] (Bauschke and Moursi, 2023) or Peaceman-Rachford operator of R_f and R_g. Simply, it is a combination of the previously mentioned two terms: $PR(f, g) := R_g R_f$. By iterating only its proximal mappings of P_f and P_g, it does not yield a minimizer of the coupled non-smooth convex function f+g. The reflectants or reflected proximal mappings are non-expansive mappings, so their composition of $R_g R_f$ as well [17] (Bauschke and Moursi, 2023)

  .

- *Douglas-Rachford splitting* - it is a combination of reflective and proximal mapping operators. More precisely, reflectant of R_f and proximal mappings of P_f and P_g, where iterating forward gradient steps or proximal steps it goes directly to minimizers of each function, so that $DR_(f, g) := I_d - P_f + P_g R_f$[17] (Bauschke and Moursi, 2023). The definition of Proximal Point Algorithm [17] (Bauschke and Moursi, 2023) actually reflects on convergence property of the function. Let f be a convex, lower semi-continuous and proper on H (Hilbert space). Assume that the set of minimizers of f is nonempty, and let $x_0 \in H$, then the sequence generated by $x_{t+1,i} := P_f(x_t)$ and converges to a minimizer of f. The same holds for function g. The dual function of f_t and g_t is closed convex and fully separable. The fix points of DR: $FixDR = FixR_g \cdot R_f$. By the firmly non-expansive definition holds that $DR = \frac{1}{2}(I + R_g \cdot R_f)$. In this context, the mentioned mathematical technique was applied in financial application of a cost minimization problem. It serves as a useful tool for optimizing the selling and buying transaction of a financial instrument (i). However, in the algorithm splitting structure, we are only speaking about neighbour terms of x_t, i: backward in time ($x_{t-1,i}$) and forward in time ($x_{t+1,i}$). Therefore, the agent is able to create a financial strategy on a short-run based on the near past experience. The purpose of the agent is to optimize implementation by minimizing transaction costs and risk.

In empirical finance, the investor tends to avoid excessive transaction cost and risk. The goal in the research is to minimize the representative dual financial function: (1.) Financial risk of the Bid-Ask Spread (f_t) and (2.) Bid-Ask Spread ($g_t(x_t - x_{t-1})$). Let $(x_t)_i \in \mathbb{R}^n$ denotes the amount of asset holdings in time period t = 1, ... T., such that x_1,...,x_T, where x_0 is the initial amount of the portfolio. The financial investor is rational and tends to minimize his financial risk by utilizing arbitrage condition on the stock market [93] (Parikh and Boyd, 2013). Reformulate, the general form of a priori estimation by Parikh and Boyd on the Standard Douglas-Rachford optimal minimum searching algorithm, we can get a domain for a two smooth convex functions. The function f is a minimization of local objective for x_i global variables, where x_0, x_1, x_2 ... x_n. The initial portfolio is x_0. The reason why was implemented the approach of a two function form is a separation of the type of financial expenses, which both has an aim to minimize costs. The two functions are a risk-adjusted negative return (f_t) in period time t; and a transaction cost (g_t) for each financial asset $x_t \geq 0$. The goal is to find, the final position to be zero, such as $x_T = 0$ or a normalization of a financial budget:

$$\text{dom }(f) = \{x \in \mathbb{R}^n \mid f(x) < \infty\} \tag{3.11}$$

$$\text{dom }(g) = \{x \in \mathbb{R}^n \mid g(x) < \infty\} \tag{3.12}$$

where, f_t is a risk-adjusted negative return in period time t; and g_t is a transaction cost for the selected financial asset $x_t \geq 0$. The higher negative return captures greater risk, while a smaller negative return indicates a mitigated risk.

The following structure of two domains basically relies on the so-called *Infimal convolution* definition as dom $g +$ dom f:

$$\min \sum_{t=1}^{T} f_t(x_t) + \sum_{t=1}^{T} g_t(x_t - x_{t-1}) \tag{3.13}$$

where, $x_t - x_{t-1}$ denotes the risk-adjusted Bid-Ask Spread moving or change in time related to the portfolio.

Theorem 3.5.1 (Moreau-Yosida envelope theorem)**.** Let f be convex, lower semi-continuous, and proper on X, and let $\lambda > 0$ [17] (Bauschke and Moursi, 2023), by the reformulation on the current portfolio example, we can get:

$$\text{env}_\lambda\, f(x_{t-1}) = f(\text{prox}_{\lambda f}(x_{t-1}) + \frac{1}{2} \parallel x_t - \text{prox}_{\lambda f}(x_{t-1}) \parallel^2 \tag{3.14}$$

According to the Moreau-Yosida envelop theorem and the proximal mapping, the envelope $\text{env}_\lambda\, f$ is $(\frac{1}{\lambda})$-smooth and its gradient $(\bigtriangledown env_\lambda f(x_{t-1}))$ is therefore $\frac{1}{\lambda}$-Lipschitz. It is $\frac{1}{\lambda}$, because we consider the Moreau envelope of f with the parameter λ in a form of infimal convolution $(f_t(x_t) + \frac{1}{2\lambda} \parallel x_t - x_{t-1} \parallel^2)$:

$$\bigtriangledown env_\lambda f(x_{t-1}) = \frac{x_t - \text{prox}_{\lambda f}(x_{t-1})}{\lambda} \tag{3.15}$$

By the implementation of the Moreau-Yosida envelop theorem, we can minimize the previous functions:

$$\begin{aligned} \operatorname{env}_\lambda f(x_{t-1}) &= \min \ f_t(x_t) + \frac{1}{2\lambda} \parallel x_t - x_{t-1} \parallel^2 \\ \operatorname{prox}_{\lambda f}\ (x_{t-1}) &= \arg\min_x f_t(x_t) + \frac{1}{2\lambda} \parallel x_t - x_{t-1} \parallel^2 \end{aligned} \tag{3.16}$$

$$\begin{aligned} \operatorname{env}_\lambda g(x_{t-1}) &= \min \ g_t(x_t - x_{t-1}) + \frac{1}{2\lambda} \parallel x_t - x_{t-1} \parallel^2 \\ \operatorname{prox}_{\lambda g}\ (x_{t-1}) &= \arg\min_x g_t(x_t - x_{t-1}) + \frac{1}{2\lambda} \parallel x_t - x_{t-1} \parallel^2 \end{aligned} \tag{3.17}$$

By solving a portfolio (x_i) optimization problem, the goal is optimizing the schedule of buying and selling the financial instrument (x_i):

$$\min \sum_{t=1}^{T} f_{t,i}(x_{t,i}) + \frac{1}{2}\lambda \parallel x_{t,i} - x_{t-1,i} \parallel_2^2 \mapsto J_\lambda^f \tag{3.18}$$

$$\min \sum_{t=1}^{T} g_{t,i}(x_{t,i} - x_{t-1,i}) + \frac{1}{2}\lambda \parallel x_{t,i} - x_{t-1,i} \parallel_2^2 \mapsto J_\lambda^g \tag{3.19}$$

$x_t \in \mathbb{R}^n$ means the portfolio of n assets (total assets) related to time period t with t $=$ 1, 2 ... T., while $(x_t)_i$ indicates only a specific asset i held by an agent in time t.

Corollary 3.5.2. The commutation property of the resolvent operator $J^f_{\frac{t}{n}}$ and $J^g_{\frac{t}{n}}$ is the following:

$$J^f_{\frac{t}{n}} \circ J^g_{\frac{t}{n}}(x) \neq J^g_{\frac{t}{n}} \circ J^f_{\frac{t}{n}}. \tag{3.20}$$

On the other hand, by the implementation of Lie-Trotter-Kato formula [90] (Ohta and Pálfia, 2017) we can get:

$$\lim_{n\to\infty} \left(J^f_{\frac{t}{n}} \circ J^g_{\frac{t}{n}}\right)^n (x) = \lim_{n\to\infty} \left(J^{f+g}_{\frac{t}{n}}\right)^n (x) \tag{3.21}$$

$$\lim_{n\to\infty} (J^g_{\frac{t}{n}} \circ J^f_{\frac{t}{n}})^n(x) = \lim_{n\to\infty} (J^{g+f}_{\frac{t}{n}})^n(x) \tag{3.22}$$

Rewriting the case on the operator D and F:

$$\lim_{n\to\infty}(e^{-F\frac{t}{n}}\,e^{-D\frac{t}{n}})^n = \lim_{n\to\infty}(e^{-(D+F)\frac{t}{n}})^n = e^{-(D+F)}t,$$
$$\lim_{n\to\infty}(e^{-D\frac{t}{n}}\,e^{-F\frac{t}{n}})^n = \lim_{n\to\infty}(e^{-(F+D)\frac{t}{n}})^n = e^{-(F+D)}t,$$
$$\log\left(\left(1+D\frac{t}{n}\right)^{-n}\right) = -n\log\left(1+D\frac{t}{n}\right) = -n\left(D\frac{t}{n}\right) - nO\left(D\frac{t}{n}\right)^2 = -Dt + O\left(\frac{1}{n}\right),$$
$$\lim_{n\to\infty}\log\left(1+D\frac{t}{n}\right)^{-n} = \lim_{n\to\infty}e^{-Dt} = -Dt,$$
$$. \tag{3.23}$$

Similarly, for the F operator, we can get the value:

$$\log\left(\left(1+F\frac{t}{n}\right)^{-n}\right) = -n\log(1+F\frac{t}{n}) = -n(F\frac{t}{n}) + O\left(F\frac{t}{n}\right)^2(-n) = -Ft + O(\frac{t}{n})$$
$$\lim_{n\to\infty}\log\left(1+F\frac{t}{n}\right)^{-n} = \lim_{n\to\infty}e^{-Ft} = -Ft$$
$$. \tag{3.24}$$

The result of operator D and F indicates that there is a downward slope of Douglas-Rachford operators in the analytical approach.

Without loss of generality, the local necessary conditions for the problem, where the objective function is a minimization problem of $f(x)$ subject to $G(x) \leq 0$. The function f(x) can be defined on a vector space X and G is a mapping from X into the normed space Z with positive cone P (D.G. Luenberger, 1969). The inequality constraints of Kuhn-Tucker theorem is closely related to the global Lagrange multiplier theorem [75] (Luenberger, 1969).

Definition 3.5.1 (Kuhn-Tucker Theorem [75] (Luenberger, 1969))**.** Let X be a vector space and let Z be a normed space with a positive cone P having non-empty interior. Let G be a mapping $G : x \to Z$. A point $x_0 \in X$ is said to be regular point of the inequality $G(x) \leq G(x_0) \leq 0$ and there is an $h \in X$ such that $G(x_0) + \delta G(x_0; h) < 0$.

Theorem 3.5.3 (Generalized Kuhn-Tucker Theorem [75] (Luenberger, 1969))**.** Let X be a vector space and Z a normed space having a positive cone P. Assume that P contains an interior point. Let $f(x)$ be a Gateaux differentiable real-valued functional on X and G a Gateaux differentiable mapping from X into Z. Assume that the Gateaux differentials are linear in their increments and x_0 minimizes $f(x)$ subject to $G(x) \leq 0$ and that x_0 is a regular point of the inequality $G(x) \leq \theta$. Then there is a $z_0^* \in Z^*$, $z_0^* \geq \theta$, such that the Lagrangian:

$$f(x) + \langle G(x), z_0^* \rangle$$

where, $\langle G(x), z_0^* \rangle = 0$. The sets A and B are convex, actually both are convex cones [75] (Luenberger, 1969).

$$A = \{(r, z) : r \geq \delta f(x_0; h), z \geq G(x_0) + \delta G(x_0; h)\}.$$

and

$$B = \{(r, z) : r \leq 0, z \leq \theta\}.$$

for some $h \in X$.

3.6 Minimization of Bid-Ask Spread

The financial transaction cost can appear in the following forms: brokerage fee, government fee, trade related costs and bid-ask spread. The latter represents a deviation between the highest asset price, which the investor would pay and the lowest price, which would the seller accept on the financial market. The assumption is that there is asymmetric information between the market participants with two types of price effect: (1.) *permanent price effect* and (2.) *temporary price effect*. The research deals with a permanent price effect.

In general, a spread is the difference between the price, interest rate or yield of two assets. One type is the bid-ask spread, which describes the difference between the bid and the ask price and can be described primarily as a function of liquidity, trading volume and volatility, while depending on the costs of asymmetric information availability, processing and inventory. Biais et al. [21] (Biais et al., 2005) in their portfolio trading game emphasis asymmetric information, where a priori and a posteriori information affect investment decision. Milgrom and Stokey [78] (Milgrom and Stokey, 1982) said in their no-trade theorem that under the condition of transparent prices, there would be the same utility function with no trading potentials, such as no speculative trade. The asymmetric information environment was also studied by Löfgren, Persson and Weibull (2002) indicated that not sufficiently informed agents can extract information for future action from those who are better informed, among the market participants [74]. The research considers that for the investor, a reliable information of the financial market is important in order to diminish financial expenses.

The Bid-Ask Spread concept can be seen through the lens of the following two properties, except the third, since it is a government or policy related expense, and it is a fixed cost:

- *Information costs* - the investor gets partial information from the market: related to the timing issuance of a portfolio and asset price. If we consider the difference between the data and information, the data is only a fact without making any action, but financial information leads to taking action. A bullish signal occurs, when a standard price of a portfolio is arbitrary chosen, for instance 290 dollar, then the financial investor gets an information that the value of portfolio is actually 500 dollar, then he would definitely realize a positive return. In the case of bearish signal, if a standard price of a portfolio is 290 dollar, than the financial investor gets an information, the value of that particular portfolio is actually 190 dollar, which would be a negative realization, a risk-adjusted negative return. By portfolio optimization the aim is to minimize information costs. Through Bayesian inference there would be a range of possible gains of trade. According to Central Limit Theorem, we suppose that the number of possible market transactions within a trading day are between P$(15 < X < 20)$ for our investor, using a binomial distribution (Bernoulli-experiment) for the probability we can get 0.79, so it is likely that the agent performs transactions between 15 and 20 in one trading day.

- *Bargaining costs* - The arbitrage opportunity enables for the investor to gain profit from a temporary price changing effect of a portfolio. Therefore, the utility function differs among the market participants.

- *Government and legal costs* - The third approach refers to the payment of a government fee or legal contract related obligations.

The financial magnitude of first two approach, where exists a Bid-Ask Spread, usually do not deter the agents from moving capital of a bank account into an asset or a financial instrument.

3.7 Minimization of Risk-adjusted negative return

The negative return incorporates financial risk factor, which can also be used to reflect the price of a stock (Figure 3.6) as bid and ask price of an assets. If the price of an equity drops below the purchased price instead of going up, it is a negative return realization. Let an investor buys a stock for 5.50 dollar/share and keeps it, till it actually starts to decrease in value, if the stock does not guarantee a payment of a dividend then the investor realizes negative return (Figure 3.7).

The assumption is that the trading risk is not the same in different time periods, moving from period x_t to $x_t - 1$, that it has a dynamic effect. A priori estimation of Biais et al. [21] (Biais, 2005) is a very good reflection on a typical trading strategy, where the agent seeks for a profit realization, such as traders, who observe the value of "not 50", they offer to buy at $x \leq 240$ (bullish signal), the traders who observe the value of "not 240" they stay out of the market, and finally the agents, who observe the

Figure 3.6: Bid and Ask Share Prices of Bissaloy Steel Australian company in AUD.

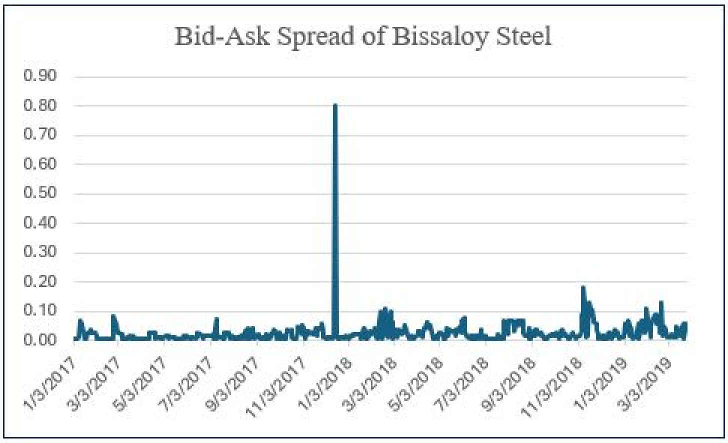

Figure 3.7: Bid-Ask Spread of Bissaloy Steel Australian company in AUD.

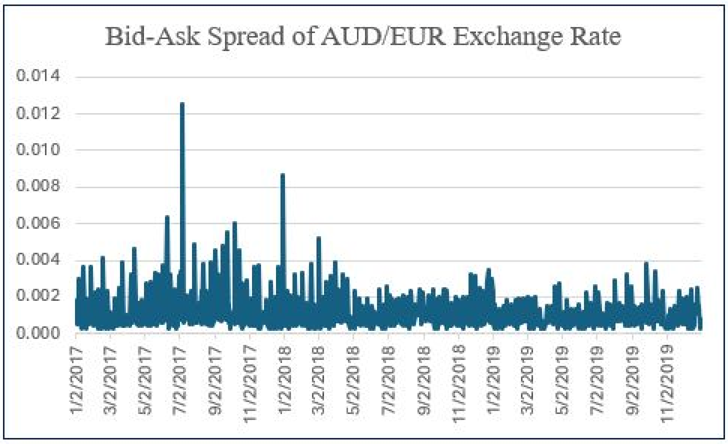

Figure 3.8: Bid-Ask Spread of AUD/EUR Exchange Rate.

value of "not 490" (bearish signal), they offer to sell at prices $x \geq 240$. The agent can regret investment if he paid too much for a not valuable asset. According to the previous authors [21] (Biais, 2005), the portfolio investor can be risk-averse, if follows the rule of overconfidence judgement, self-monitoring and trade performance on the market.

3.8 Numerical examples

The Standard Douglas-Rachford iterative algorithm solves convex optimization problem, when minimizing dual functions of $f(x)+g(x)$. The author, E. Bertsekas [20] (Bertsekas, 1989) in the journal of Mathematical Programming was elaborating about the iterative method of Douglas Rachford operator splitting algorithm. If we reformulate the Standard Douglas Rachford on our economic application of f and g convex functions, we can get the following *Douglas Rachford Iteration Algorithm*[76] (Mahey and Lenoir, 2017), where the initial point or initial portfolio of Bid-Ask Spread for a Company Share is x_0 and let's introduce a Bid-Ask Spread for Foreign Currency Exchange Rate y_0, t=0 and $\lambda > 0$.

3.8.0 Bid-Ask Spread of a Company Share Prices and Foreign Currency Exchange Rate

Bid-Ask Spread of a Share Price is calculated for an Australian company, Bissalloy Steel, where the daily trading data are denominated in Australian dollar (Figure 3.7). In relation with this, the AUD/EUR exchange rate Bid-Ask Spread (Figure 3.10) is a subject of examination, where there is a question that for 1 AUD how much EUR needs to be paid by an investor? In comparison analysis, the magnitude of Bid-Ask Spread of the exchange rate is smaller than the Bid-Ask Spread of an issued share by the Australian company (Figure 3.7 and 3.8.).

The Douglas Rachford Iteration Algorithm consists of the following steps:

1. $x^{t+\frac{1}{3}} = J_\lambda^f(x^t + \lambda y^t)$
2. $y^{t+\frac{1}{3}} = \lambda^{-1}(x^t - x^{1+\frac{1}{3}} + y^t)$
3. $x^{t+\frac{2}{3}} = J_\lambda^g(x^{t+\frac{1}{3}} - \lambda y^{1+\frac{1}{3}})$
4. $y^{t+\frac{2}{3}} = \lambda^{-1}(x^{t+\frac{2}{3}} - x^{1+\frac{1}{3}}) + y^{t+\frac{1}{3}}$
5. $(x^{t+1}, y^{t+1}) = \frac{1}{2}((x^{t+\frac{2}{3}}, y^{t+\frac{2}{3}}) + (x^t, y^t))$
6. $t+1 \mapsto t;\ \ t \mapsto t-1$ etc.

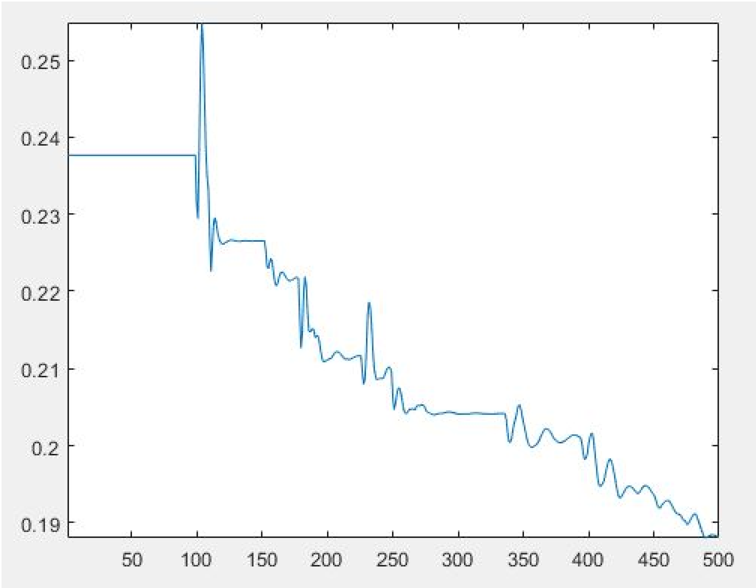

Figure 3.9: Douglas Rachford convergence of Bissalloy Steel Bid-Ask Spread for 500 iterations

3.8.1 Empirical Data

Data frequency is in trading days (average 252 days per year). The 10-Year Italian Generic Government yields consist of Net Yields after taxes denominated in EUR. The yields are based on the bid side of the market and updated each trading day for the period between: 01.01.2020-22.03.2024. Data source: Bloomberg (Ticker: GBTP10YR). The second financial instrument represents a realization of a negative return, which we want to minimize in the model period by period using the splitting algorithm, where the Italian share, Terna-Rete Elettrica Naziona SpA (transfers high-voltage electricity and grid in Italy) is a riskier equity than the Italian government bond. Through its subsidiaries, it covers a substantial proportion of the national electricity transmission grid. Data source of Terna-Rete Elettrica Naziona daily trading asset return in EUR: Bloomberg (Ticker: TRN IM Equity) for the period between: 18.06.2010.-20.03.2020. The third financial instrument illustrates a Bid-Ask Spread for an Australian share, Bissalloy Steel Group Ltd. for the period between 03.01.2017-25.03.2019. Data source: Bloomberg (Ticker: BIS AU Equity), the time series data only was available till 25.03.2019. The Bid-Ask Spread denotes a difference between the Ask price (the minimum price, which the portfolio seller is willing to take) and the Bid price (the maximum price, which the portfolio acquisitor is willing to give). By definition, it is a cost in association with the financial instrument, such as equity and foreign currency exchange rate. In order to calculate the Bid-Ask Spread, we subtract from Ask price the Bid price: Bid-Ask Spread = Ask price - Bid price. It can fluctuate based on the supply and demand of security. Another form of a Bid-Ask Spread is related to the exchange rate, where the Australian dollar and Euro were taken as a Bid-Ask Spread for the

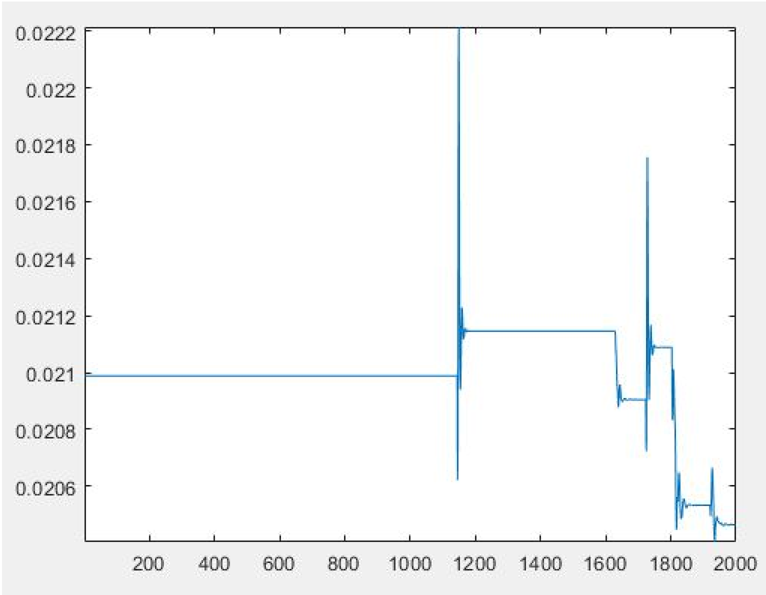

Figure 3.10: Douglas Rachford convergence of AUD/EUR Foreign Currency Bid-Ask Spread for 2000 iterations

period between: 02.01.2017-31.12.2019. Data source: Bloomberg (Ticker: AUDEUR BGN Curncy). The question is how much Euro has to be paid for one unit of Australian dollar? The result indicates that the Bid-Ask Spread is very low in comparison with the higher Bid-Ask Spread of Company's share.

3.9 Conclusion on Douglas Rachford Operator Splitting Algorithm

Ardia, D., Guidotti, E., Kroencke, T.A. [8] (Ardia et al., 2024) in their work of *Efficient Estimation of Bid-Ask Spreads from Open, High, Low, and Close Prices* implemented the estimator in a new programming language on data, which contains monthly estimates of the effective bid-ask spread for each stock in the CRSP (Center for Research in Security Prices) of the U.S. Stock. Ardia, D., Guidotti, E., Kroencke, T.A. stated three assumptions:

1. **Assumption 1**: Fundamental returns are not serially correlated;
2. **Assumption 2**: Fundamental returns are uncorrelated with bid-ask bounces;
3. **Assumption 3**: Bid-ask bounces are uncorrelated and have zero mean.

However, Ardia, D., Guidotti, E., Kroencke, T.A. used a high frequency data, where their simulations used a constant spread of 1%, and the observed a trade ranges are from 1/390 to 1, such that the expected number of daily trades ranges from 1 to 390 [8].

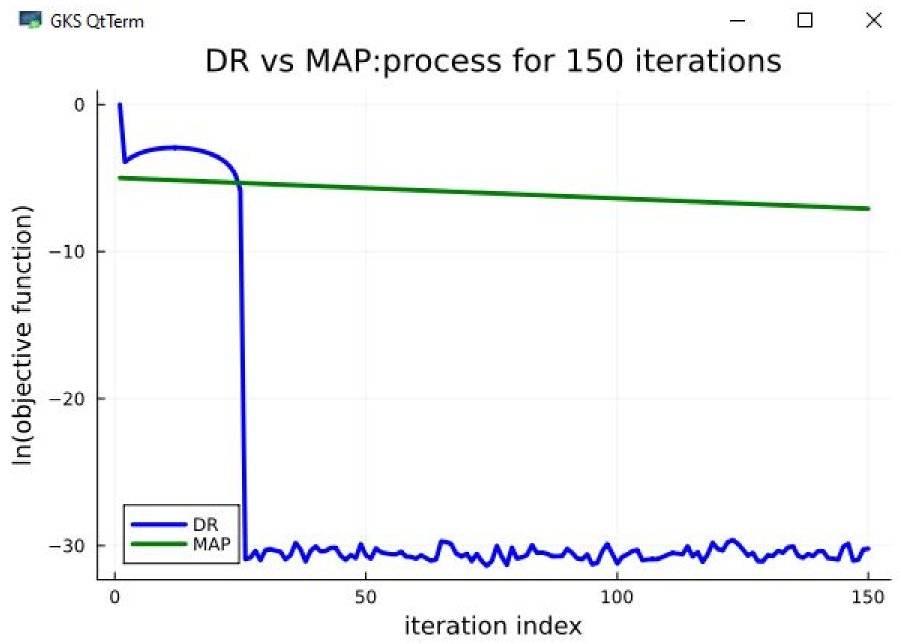

Figure 3.11: Douglas Rachford vs Method of Alternating Projections

The advantage of Douglas Rachford Algorithm is a useful splitting technique for finding a zero (minimum, the lowest value) of the sum of two maximally monotone operators such as of two financial function (f_t of risk adjusted negative return and g_t of transaction cost). The disadvantage of Douglas Rachford Algorithm is when the sum problem has no zeros.

The Douglas Rachford Iteration Algorithm can be executed in Julia and Matlab. The algorithm minimizes f(x) + g(x) and solves this convex optimization problem. It is a subject of numerical iteration as it was previously introduced, the iteration steps in Part 7. The more general form of splitting algorithm, the ADMM is possible to execute in Python [93] (Parikh and Boyd, 2013). The multi-period portfolio optimization can be rewritten in matrix form, where each period is a sum of each type of costs; and summing up all the periods, we can get the total cost of X and Y:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_T \\ y_1 & y_2 & \dots & y_T \end{pmatrix} \in \mathbb{R}^n$$

The matrix denotes the total sequence of each financial investment. The agent runs long security positions in the expectation that the value of a stock will rise in the future. (The opposite would be a short position, where the investor, who sells short position believe that the price of the value of the stock will decrease in the upcoming future). The total cost for each portfolio in parallel considering a Bid-Ask Spread for a company share (X) with high-risk of cost loss and Foreign Currency Exchange Rate with low-risk of cost loss(Y):

$$f(X) = \min \sum_{t=1}^{T} f_{t,i}(x_{t,i}) \tag{3.25}$$

$$g(X) = \min \sum_{t=1}^{T} g_{t,i}(x_{t,i} - x_{t-1,i}) \tag{3.26}$$

and,

$$m(Y) = \min \sum_{t=1}^{T} f_{t,i}(y_{t,i}) \tag{3.27}$$

$$z(Y) = \min \sum_{t=1}^{T} g_{t,i}(y_{t,i} - y_{t-1,i}) \tag{3.28}$$

The agent invests in the second period. Under the condition of high- and low-risk circumstances, the investor has a primary focus on the decrease of the Bid-Ask Spread. We can observe that in both cases, when there is a high-risk of a company's Bid-Ask Spread and a low-risk of the agent's trading speculation into Foreign Currency transactions, the optimal minimization can be reached by analytical and numerical solutions. Julia software considered a total of 150 iterations (Figure 12.), and was compared to 200 iterations [17] (Bauschke and Moursi, 2023) in the first round (A). In the second round (B), the following set up of codes was implemented in Matlab with downloading additional packages of Toolbox signal and Toolbox general [95] (Peyré, 2014) for Standard Douglas-Rachford iterative algorithm reformulated on operator F and D.
Data were extracted from Bloomberg Terminal and it is available in the following github repository: Douglas-Rachford Operator Splitting Algorithm. Details regarding the algorithmic trading strategies in the context of Buy Low and Sell High can be read in the work of [43].

CHAPTER 4

Quantitative Tightening: Theory, Research, and Impact on Selected Emerging Market Economies

4.0 Introduction

Chapter 4. describes the Quantitative Tightening approach - Theory, Research, and Impact on Selected Emerging Market Economies [1]. The world economies were in the process of recovering during and after the pandemic and inflation exceeded its target level; most of the central banks used tightening policy by reducing the central bank balance sheet size, i.e., Quantitative Tightening (QT). The Federal Reserve faced again with a question: How to manage its balance sheet? Beside the process of Fed's balance sheet shrinkage, Gál and Juhász (2025) follow the long-end yield change and their spillover effect from a hegemon country (USA) towards the emerging market countries. The balance sheet shrinkage policy became relevant as it supports the conventional monetary policy of avoiding further inflationary pressures [42].

The World Bank authors (Lim, Mohapatra, and Stocker 2014) evidented that the unconventional monetary policy in the form of Quantitative Easing affects gross financial inflow from the United States to the other advanced and developing countries, where the government bond flow channel tends to be even more sensitive, than other foreign direct investments [71]. We saw that these effects of quantitative tightening are visible

[1]See as a reference: Gál, H., Juhász, A. *Quantitative Tightening: Theory, Research, and Impact on Emerging Market Economies*, Journal of Central Banking Theory and Practice, 1 (2025), pp. 163-181. https://doi.org/10.2478/jcbtp-2025-0009.

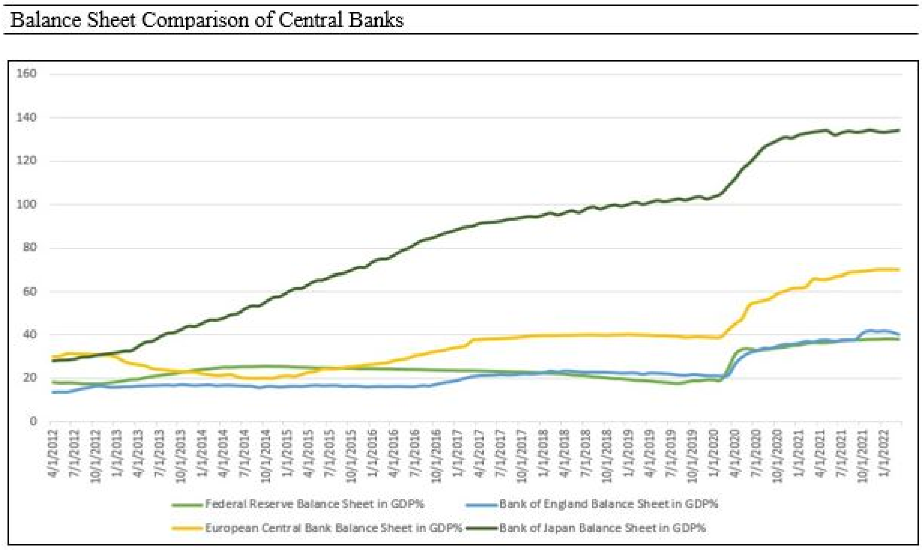

Figure 4.1: Balance Sheet Comparison of Central Banks. *Source: Bloomberg*

in the changes of interest rates and asset prices.

A number of economies in the world have raised their sovereign debt level during the Covid-19 pandemic. In order to cover their accelerated budget deficit, the governments issued more securities, where in the role of investors came up not only the traditional public or the private sector, but also the central banks. The Federal Reserve (Fed) without any obligation or government forces has the autonomy to purchase securities, this occurred during the Covid-19, where the Fed consistently by the fourth round of Quantitative Easing prevented nominal yields from turning negative (Allen, D. Kyle and Hein E. Scott, 2023) [67]. Figure 4.1. shows that it initiated the asset purchase program, which led to a significant expansion of Fed's Balance Sheet and illustrates the balance sheet comparison of the four major central banks for nearly a decade period. Actually, if we see, it starts from a moderate level and than increases gradually. It is visible that the liquidity excess is much higher in 2022 than in 2012. The balance sheet level of the Fed and the Bank of England in GDP % slowly follows each other. While, the Bank of Japan's acceleration was outstanding compared to its peers! The European Central Bank with higher volume follows similar periodical jumps such as the Fed and Bank of England.

During the COVID-19, the Fed announced $700 billion purchase of assets on March 15, 2020 [2]: $500 billion of Treasury Securities and $200 billion of Mortgage-Backed Securities (MBS) to support liquidity needs of the financial market in response to the pandemic [3]. The accumulation of securities more than doubled on the Fed's balance

[2]Federal Reserve issued FOMC (Federal Open Market Committee) statement on March 15,2020, related to the asset purchase program. Source: Board of Governors of the Federal Reserve System.

[3]The FOMC purchased Treasury securities and agency Mortgage-Backed Securities in the optimal amounts needed to support smooth market functioning and effective monetary policy transmission. Source: Federal Reserve announces extensive new measures to support the economy Board of Governors

sheet compared to the period before the pandemic. It increased from $4 trillion to 9$ trillion of Treasuries and MBS. At that time, it was not possible to predict how much the Fed's balance sheet would increase? By the end of 2020, the Fed's balance sheet was standing at 34% of GDP in the U.S., while the ECB's at 59% in Europe, the Bank of England's 40%, and the Bank of Japan's 127%. The Fed's policymakers observed that the balance sheet is higher than the average level, and further shrinkage is necessary to happen in 2022! The Fed's asset purchase program ended in March 2022 [4]. The list of same open market purchase transaction according to the International Monetary Fund was utilized by the large number of other countries from the advanced and emerging market economies [5]. The Fed's first intention was to increase its primary policy rate [6] in order to fight against the inflation, which was increasing [7]. The authors (Lim, Mohapatra, and Stocker 2014) from the World Bank found that 62% of the increase in financial inflows was during 2009–2013 from the US to other advanced and developing countries, where at least 13% of this was attributed to QE in high-income countries and 5% for the average developing countries [71]. We estimated that the monetary policy shock from the U.S. raises long-term yields by 42.70 basis points related to the top five major emerging markets and 85.32 basis points for other selected emerging markets.

4.1 Balance Sheet Shrinkage Policy

Quantitative Tightening or Quantitative Tapering is a process of decreasing the fixed-income assets held by the Fed. This paper differentiates the active and passive approaches. First is a shrinkage of the balance sheet by actively selling securities. The passive approach means that the securities on the balance sheet expire automatically without any financial transaction. In the last phase, when treasury securities reach their maturity date, they are paid off by the government and Mortgage-backed securities are paid off by government-sponsored enterprises, such as the Federal National Mortgage

of the Federal Reserve System.

[4] The same situation occurred in Europe, the European Central Bank (ECB) announced the discontinuation of its Pandemic Emergency Purchase Program (PEPP) at the end of March 2022.

[5] List of Asset Purchase Programs for Advanced economies: Australia, Canada, Iceland, Israel, South Korea, New Zealand, Norway, Sweden, United Kingdom. Developed and Emerging Market Economies: Angola, Bolivia, Cabo Verde, Chile, China, Colombia, Costa Rica, Croatia, Egypt, Hungary, India, Indonesia, Jamaica, Mauritius, Mexico, Philippines, Poland, Romania, South Africa, Thailand, Turkey, Ethiopia, Ghana, Papua New Guinea, Rwanda, Uganda. List of Asset Purchase Programs, p.27-31. In IMF Working Paper written by Chiara Fratto, Brendan Harnoys Vannier, Borislava Mircheva, David de Padua, and Hélène Poirson, January 2021.

[6] FOMC Communications Related to Policy Normalization of plans and principles for reducing the size of the Federal Reserve's Balance Sheet. Source: Board of Governors of the Federal Reserve System.

[7] Ha, Kose, and Ohnsorge (2022) from World Bank compared similarities and differences between the Great Inflation in the 1970s and the Inflation of 2022. Similarities appeared in supply disruptions driven by the pandemic, supply shocks of energy prices resemble the oil shocks in 1973 and 1979–1980 and heightened geopolitical tensions.

Assets		Liabilities	
U.S. Treasury securities	5,700	Reserves (*depository institutions*)	3,271
MBS	2,726	Repurchase agreements	2,494
Other	425	Currency	2,276
		Treasury general account	530
		Other	280
Total	**8,851**	**Total**	**8,851**

Table 4.1: Fed's Balance Sheet Composition. *Data source: Federal Reserve*

Association or the Federal Home Loan Mortgage Corporation. The QT1 lasted for less than two years, between 2017 and 2019. Essentially, this tightening policy had never been done before on a massive scale in the United States. Engemann (2019) defines it as redeeming or reducing of the Fed's balance sheet size [37]. However, Hopper (2018) reflects on it as the normalization of monetary policy through "unwinding," where the term "unwinding" means the slow and gradual nature of the reduction of the Fed's balance sheet, which was expanded after the global financial crisis [68]. Wheelock (2018) interpreted "unwinding" as simply stopping the replacement of securities, which mature. Waller (2018) pointed out that it is a process of increasing the supply of Treasuries in the financial market, where the Fed explicitly lets the supply of Treasuries in the hands of the private sector grow. The tightening process of selling securities causes an increase in supply, which pushes the prices low, and the yields are adjusted upwards. The selling of treasury securities on the asset side of the balance sheet is decreasing, while the liability side of the balance sheet exactly matches the decrease in assets (Hollenhorst et al.,2022). The Fed's balance sheet represents equilibrium on both the assets and liabilities side (Table 4.1.):

The Federal Open Market Committee (FOMC) members implemented QT2 in June 1, 2022. The main reasons behind the voting for quantitative tightening program:

1. the inflation is much higher than the average level and this complementary tightening operation could help to fight against inflation. The price increased in all sectors, in the supply chain, energy, fuel and non-durable goods;

2. over-accumulation of assets on the Fed's balance sheet also urged the quantitative tightening operation.

Starting the QT2 was subject to being over the zero lower bound. Therefore, the Federal Funds Target Rate was increased from 0.08% (range 0.00-0.25) to 0.33% (range 0.25-0.50) in mid-March, 2022; then from 0.33% (range 0.25-0.50) to 0.83% (range 0.75-1.00) at the beginning of May, 2022. After the base rate further increased from 0.83% (range 0.75-1.00) to 1.58% (range 1.50-1.75) in mid-June, 2022. Guillermo Peña (2023) stated in his research that raising interest rates mainly affected middle-income households in Europe during the Covid-19 and caused asymmetric repercussions in the level of income [94].

4.2 Literature review

Monetary policy has gone through various changes over the last decades, which witnessed the end of the "reserve position doctrine" and the turn to a clear focus on short-term and long-term interest rates. The reserve position doctrine was supported by Keynes and later by the monetarist school, developed mainly by the U.S. central bankers during the early 1920s. It became the unchallenged principle for over sixty years. Bindseil (2003) reflects in his book on monetary policy history and explains the three supporting instruments of monetary policy: open market operations, standing facilities, and reserve requirements, where open market operations are the subject of QE and QT [**bindseil03**]. Five years ago, the incumbents thought that the balance sheet was a poorly understood tool according to Hollenhorst et al. (2022) and any further effect on the markets should be slowly and quietly unwound. D'Amico and Seida (2020) performed a quantitative analysis, and they found that the yield sensitivities of QT are more prominent on average, than yield sensitivities of QE [33]. While, Goodhart (2010) emphasizes that the role of the central bank through its balance sheet operations is important [50]. These financial instruments serve to reach targets, where the Fed has a dual mandate: price stability and reaching maximum employment! Over the past three decades, central banks of advanced economies have established a credible track record of achieving inflation targets (Bordo et al., 2007; Eichengreen, 2022) [23], [36]. Other authors (Cecchetti and Tucker 2021) pointed out that official declarations of aiming for price stability and financial stability are unsatisfactory if there is a jump to end goals without attending to the specific operations and facilities [29]. The authors alternatively provide a categorization of operations: (1.) stimulating or dampening aggregate demand by monetary policy to achieve price stability with full use of the economy's productive resources; (2.) lending funds to firms whose financial needs cannot be met via private markets – central bank as a lender of last resort; (3.) addressing liquidity problems in specific markets – central bank as a market maker of last resort; (4.) ensure the flow of credit to specific sectors, regions, or firms – central bank as a selective credit supporter; (5.) providing needed funds directly to the government – central bank as in an emergency government finance role, where we added the (6.) operation of balance sheet policy and its international transmission mechanism channel. According to a numerical estimation of Wei (2022), a $2.2 trillion passive roll-off of nominal treasury securities from the Federal Reserve's balance is equivalent to an increase of 29 basis points in the current federal effective funds rate at ordinary times, over the three years of observations, and 74 basis points increase during the crisis periods [105]. Similarly,

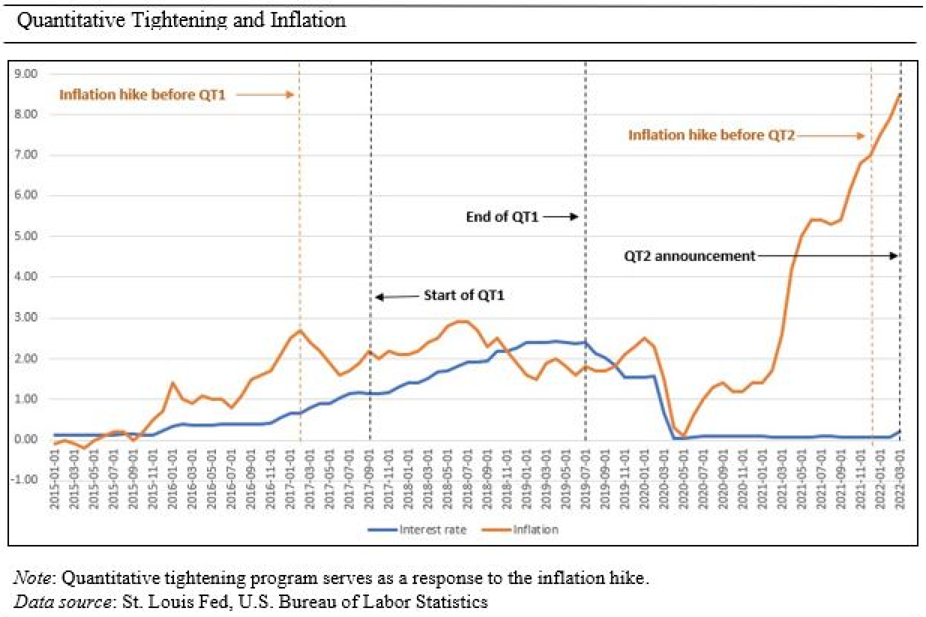

Figure 4.2: Quantitative Tightening and Inflation. *Data source: St. Louis Fed, U.S. Bureau of Labor Statistic*

Crawley et al. (2022) predict reducing the size of the balance sheet by about $2 trillion over the next few years, which would be roughly equivalent to raising the policy rate by a little more than 50 basis points [32].

4.3 Macroeconomic Outcomes: Prior vs. Current Cycle of QT

The prior cycle could serve as a reference point for the current one, where the policy rate remains the primary and active tool of the monetary policy, and the quantitative tightening program a secondary and complementary tool in the background. In comparison with QT1 program, the QT2 is expected to gear up faster, related to a more rapid runoff of short-maturity Treasuries. Figure 4.2. provides an overview of inflation behaviour through quantitative tightening programs by the Fed.

The Fed was more conservative with its balance sheet reduction in QT1. The only $10 billion/month was followed by a reduction of $50 billion/month with a split of $30 billion in Treasuries and $20 billion in Morgaged-Backed Securities (MBS). During the time, when it initiated its QT1, in 2017, the total balance sheet was around $4.5 trillion, and it managed to bring it down to $3.8 trillion. In comparison, related to the maturity structure of the treasury portfolio for QT2 is longer than in QT1. The second tapering has started with a $47.5 billion/month decrease in September of 2022 and followed by a $95 billion/month reduction of $60 billion in Treasuries and $35 billion in MBS.
Some of the features of balance sheet tightening program, which can be observed:

The maturity distribution of the selected assets (billion $)

Maturity	Within 15d	16d-90d	91d-1y	1y-5y	5y-10y	Over 10y	All
U.S. Treasury security holdings	78,233	324,425	819,487	2,017,728	1,009,321	1,451,434	5,700,628
Weekly changes	-44	+615	-505	+565	+22,583	-21,761	+1,453
MBS holdings	0	2	57	2,412	57,525	2,665,909	2,725,906
Weekly changes	0	0	0	-1	-2	-1,565	-1,567

Note: "d" stands for "days", "y" stands for "years".
Data source: Federal Reserve statistical release (H.4.1.), August 25, 2022

Table 4.2: The maturity distribution of the selected assets. *Data source: Federal Reserve statistical release (H.4.1.), August 25, 2022.*

1. The tightening of the Fed's balance sheet affect long-term yields;

2. Money supply in the system decreases as it induces less money in the market circulation;

3. Small and recursive active sales occures, when passive runoff of Treasuries and MBS reach below the monthly target. Small amount of sales is less likely to involve market concerns;

4. Inflation, debt, and income inequality could endanger recovery in emerging economies, according to the World Bank's Global Economic Prospects report [8];

5. The Fed officials can opt to reinvest toward shorter maturities, for instance, by purchasing several 3-months T-bills in the form of shorther maturity resinvestments;

6. The trade-off between balance sheet reduction and effective funds rate hike – some market participants use the „3 to 1" rule of thumb, such that the 3-basis point (bp) move in policy rates is worth 1 bp on 10-year treasury yield change. The $50 billion per month of balance sheet reduction is a substitution for 2% additional rate hikes per year. The Kansas City Fed President, Ester George, mentioned that policy rate hikes might slow once the balance sheet reduction starts (Hollenhorst et al., 2022).

Besides treasury and MBS securities, there are other assets on the Fed's balance sheet in a much smaller amount, such as FX swap lines and the Fed's discount window.

[8] Global Economic Prospects Report, Stagflation Risk Rises Amid Sharp Slowdown in Growth World Bank, 2022.

Period	Quantitative Tightening (QT)
	Prior cycle (QT1)
June 2017	Plan for balance sheet reduction announced
September 2017	The balance sheet reduction started at $10 billion/month in October
October 2018	Speed of $50 billion/month ($30 billion/month treasury bills and 20$ billion/month Mortgage-Backed Securities – MBS) reached.
January 2019	The committee announced that it would implement a policy in an "ample reserves" regime.
March 2019	Announced that balance sheet reduction will be fully phased out in September 2019
July 2019	Balance sheet reduction ended two months earlier than planned, along with a 25bp cut in the policy rate.

Table 4.3: Quantitative Tightening: 20217-2019

FX swap lines provide U.S. dollars to the five major central banks. However, given the stability in U.S. dollar funding markets (Cabana, Craig 2022), its usage is minimal.

4.3.0 Ben Bernanke: Taper Tantrum

Ben Bernanke, then-chairman of the Fed, recommended tapering asset purchases in 2013. The outcome was that the yield on 10-year U.S. Treasuries rose from around 2% in May 2013 to around 3% in December 2013. This sharp climb in yields is often referred to as the "taper tantrum". In July 2021, Fed officials signaled that the Federal Reserve would need to reduce the volume of its bond purchases. This signal made some investors worry (Bernanke, 2022) about another "taper tantrum", due to its negative connotation and shock to the market [19]. Despite these fears, most investors remained placid when the Fed hinted at tapering in July 2021. Essentially, the announcement was in line with market expectations in 2021, while the announcement in 2013 came much earlier than expected! Ben Bernanke indicated that a balance sheet reduction should follow the effective policy funds rate increase as a first phase of the tightening process, but he was taking no position on the appropriate pace of monetary tightening. The policy communication is easier, and the risk of market disruption is minimized if the shrinkage of the balance sheet is predictable. It is prudent not to begin that process until short-term interest rates are comfortably away from their effective lower bound. For instance, the first quantitative tightening program did not begin until rates had reached 1.00-1.25.

	Current cycle (QT2)
June 2022	Balance sheet reduction from June 1. Declining split between \$30 billion/month Treasuries and \$17.5 billion/month Agency debt and Agency MBS.
September 2022	Rise at monthly intervals to \$95 billion/month (split of \$60 billion/month Treasuries and 35\$ billion/month MBS).
Mid-2024	Sufficient liquidity has been drained, so the reserve repo facility (RRF) is no longer used. As a result, bank deposits may move sideways as the depletion of the RRF means reserves shrink faster.
Mid-2025	The balance sheet approaches \$6 trillion in assets (\$4 trillion in treasury holdings, \$2 trillion in MBS holdings) and \$2.5 trillion in reserves. At this point, balance sheet reduction might be slowed or stopped as Fed officials seek to maintain a comfortable level of aggregate reserves. According to estimation, the \$2.2 trillion passive roll-off of nominal treasury securities from the Federal Reserve's balance sheet over three years is equivalent to an increase of 29 basis points in the current federal funds rate.

Note: Projections shown are based on Rajdeep Sengupta and A. Lee Smith: Assessing Market Conditions ahead of Quantitative Tightening, Economic Bulletin, Federal Reserve of Kansas City, 2022. and Wei, B. 2022. "*How Many Rate Hikes Does Quantitative Tightening Equal?*", Federal Reserve of Atlanta's Policy Hub, Center for Quantitative Economic Research, No. 11–2022 July 2022. The Fed's prior versus current QT cycle, where the projections usually refer to three year period (Sengupta and Smith, 2022; Wei, 2022)

Source: Fed of Kansas City, Fed of Atlanta, Fed of New York

Table 4.4: Quantitative Tightening: 2022-2025

4.4 The general transmission mechanism of QT

Through the transmission mechanism channel of quantitative tightening program, the asset sales affect even cross-border capital flows and market yields (Gagnon et al., 2011 [41]; Hamilton and Wu, 2012 [53]; Jacob de Haan, Schoenmaker, Wierts, 2020) [52]. Market participants adjust their investment portfolios according to the operations of the Fed through the portfolio rebalancing channel. These market transactions, in turn, increase the available stock of privately held assets by buying securities. The second channel is the signaling channel (Jacob de Haan, Schoenmaker, Wierts 2020), where the central bank communicates its expectations according to the base rate and the quantitative tightening program [52]. It improves the perception of financial institutions, businesses, and households by diminishing their concerns. The third is the liquidity channel (Gagnon et al., 2011 [41]; Joyce et al., 2011 [59]; Krishnamurthy and Vissing-Jørgensen, 2013) [63], where its effects are moderate with less money in the financial circulation but the expansion of yields on securities.

The amount of money in circulation decreases in order to slow down inflation. Through the confidence channel, it affects other countries by increasing yields of securities. The unconventional monetary policy transmission channels directly affect the yield curve and price stability (Figure 4.3.).

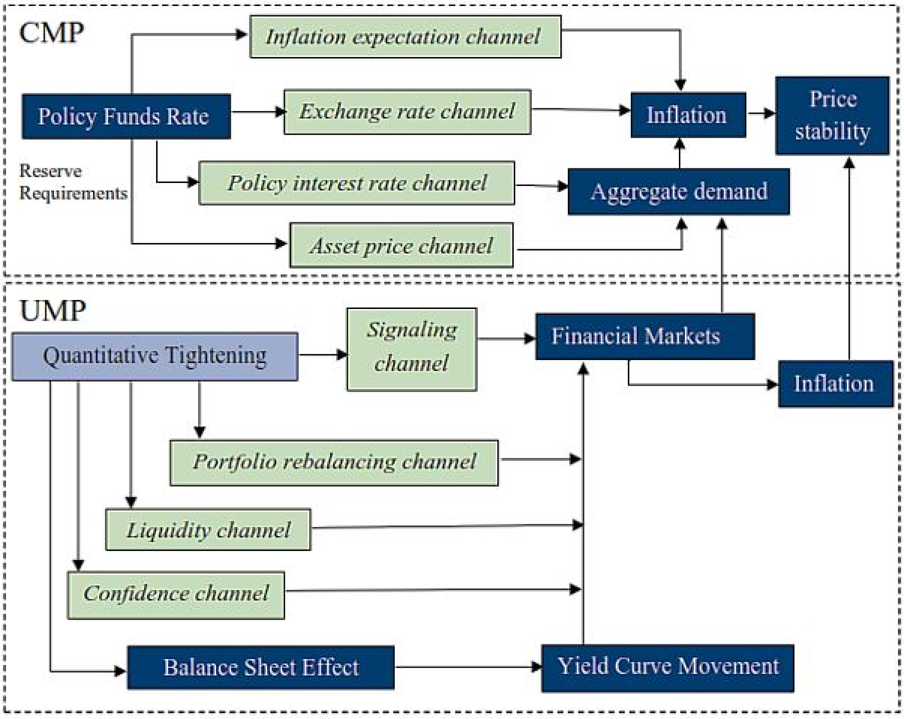

Figure 4.3: Transmission mechanism of CMP and UMP. *Source: Authors' analysis*

4.5 Economic outlook

The paper reflects on the 10-year U.S. treasury yield spillover effect towards the selected emerging market. An insightful analysis was performed by Inda Mulaahmetović (2022) related to unconventional monetary policy and macroeconomic performance in the United States [82]. The BRICS are known for their significant influence on their respective regional affairs to other neighboring countries in their region. The second group of analysis refers to an arbitrary sampling of the countries: Mexico, Peru, Hungary, Greece, and Turkey. The result shows that the yields moved upward during the QT1 operation between 2017 and 2019. The empirical result reflects on that there is significant increase in yields and a strong correlation between the rates moving across the world.

The table indicates, that the USA as a hegemon country, fulfills the (a.) high GDP, (b.) low unemployment, (c.) policy interest rate hike preliminary conditions of the implementation of the QT program.

The monthly data in local currency for each country was decomposed into five cohorts:

1. May 2012 – May 2022: a full decade overview;
2. May 2012 – June 2017: a period from 2012 till the announcement of QT1;
3. June 2017 – October 2017: a period from QT1 announcement till the QT1 program

The yield on U.S. treasury securities at 10-year constant maturity

Note: Daily yields on 10-year U.S. Treasuries between July 2017 and July 2022.
Data source: FRED (St. Louis Fed)

Figure 4.4: Quantitative Tightening and Inflation

Key Statistics of the USA and BRICS (in %)

	April 2022				2021-2022
	Inflation	**10-year Yields**	**Policy interest rate**	**GDP 2021Q4**	**Unemployment rate 2022Q1**
Brazil	12.1	12.29	11.75	1.60	11.1
India	7.8	7.14	4.25	5.40	7.6
China	2.1	2.84	1.50	4.0	5.80
Russia	17.8	10.17	17.00	5.0	4.20
South Africa	11.9	10.36	4.25	1.7	35.3
USA	**8.3**	**2.93**	**0.50**	**5.5**	**3.80**
Mean	**10.34**	**8.56**	**7.75**	**3.54**	**12.80**

Note: Mean only refers to the BRICS.
Data source: World Bank, Bloomberg (World Economic Statistics)

Table 4.5: Key Statistics of the USA and BRICS

Correlation of the USA and BRICS

Time frame/ Selected Countries	(1)	(2)	(3)	(4)	(5)
Brazil	0.46	0.23	0.09	0.72	0.60
India	0.61	0.29	0.57	0.90	0.85
China	0.59	0.70	0.34	0.27	0.04
Russia	0.39	0.07	0.58	0.29	0.62
South Africa	0.16	0.23	0.59	0.24	0.30
Mean	**0.44**	**0.30**	**0.43**	**0.48**	**0.48**

Note: The correlation of 10-year Yields between the USA and India was the highest during QT1 (Column 4).

Data source: Bloomberg (GGR - Generic Government Rates)

Table 4.6: Correlation of the USA and BRICS

Key Statistics of the Other Emerging Markets (*in %*)

	April 2022				2021-2022
	Inflation	**10-year Yields**	**Policy interest rate**	**GDP 2021Q4**	**Unemployment rate 2022Q1**
Hungary	**9.5**	6.9	5.40	7.1	3.57
Mexico	**7.7**	8.32	6.50	1.1	3.47
Greece	**10.13**	3.33	0.00	7.4	12.9
Turkey	**69.97**	8.65	14.00	9.1	11.40
Peru	**7.96**	7.85	4.50	3.2	8.97
Mean	**21.05**	**7.01**	**6.08**	**5.58**	**8.06**

Note: Selected Other Emerging Markets, where the population is lower than in BRICS.
Data source: Bloomberg (World Economic Statistics)

Table 4.7: Key Statistics of the Other Emerging Markets

Correlation of the USA and Other Selected Emerging Markets

Time frame/ Selected Countries	(1)	(2)	(3)	(4)	(5)
Hungary	0.26	0.01	0.46	0.57	0.72
Mexico	0.44	0.52	0.52	0.37	0.86
Greece	0.11	0.53	0.06	0.50	0.67
Turkey	0.01	0.64	0.90	0.01	0.40
Peru	0.43	0.15	0.13	0.63	0.66
Mean	**0.26**	**0.39**	**0.41**	**0.41**	**0.65**

Note: The mean is the highest in the fifth time frame (Column 5).
Data source: Bloomberg (GGR - Generic Government Rates)

Table 4.8: Correlation of the USA and Other Selected Emerging Markets

starts; (4.) October 2017 – July 2019: a full period of the QT1;

4. July 2019 – May 2022: a period from the end of QT1 until the current period of QT2 announcement.

Table 4.6 - Column (4) indicates that during QT1, the highest correlation was between the USA and India (0.90) and Brazil (0.72). Table 4.8 - Column (4) indicates that during QT1, the highest correlation was between the USA and Peru (0.63).

The regression indicates 10-year government bond yields with the intention to validate the hypothesis, which is that the 10-year U.S. government bond yield has a significant impact on the selected emerging market.

The change in yields in the U.S. (Table 4.10.) was an increase of 0.85%, then the change in yields in India increased by 0.99%, in Russia with a 1% annual increase in 10-year government yields. Related to the results, the variation in yields in Brazil increased by 0.22% and in South Africa by 0.29%. Only in China was a decrease in the 10-year government yields for the selected period with -0.36%.

Table 4.11. describes a change of 10-year government yields in Turkey, which was a increased of 2.19%, then the change in yields in Mexico increased by 1.61%, and a

Dependent variable: 10-year government yields (EM country)					
	Independent variable (10-year U.S. government yield)	**t-statistics**	R^2	**F-statistic**	**Intercept**
Brazil	2.21*** (0.19)	11.61***	0.77 (1.10)	128.61***	30.31*** (5.88)
India	0.50*** (0.08)	6.68***	0.77 (0.42)	127.67***	12.03*** (1.71)
China	0.37*** (0.05)	7.39***	0.55 (0.31)	47.45***	-0.54 (0.15)
Russia	0.97*** (0.21)	4.53***	0.15 (1.46)	20.52***	6.14*** (0.46)
South Africa	0.19* (0.10)	1.95*	0.66 (0.48)	78.37***	7.98*** (2.70)
Hungary	0.67*** (0.23)	2.94***	0.07 (1.55)	8.64***	2.45*** (0.48)
Mexico	0.54*** (0.15)	3.68***	0.35 (0.76)	20.39***	6.80 (4.63)
Greece	2.39** (0.96)	2.48**	0.18 (4.94)	8.21***	2.71 (43.0)
Turkey	1.08*** (0.11)	9.47***	0.78 (0.61)	137.08***	45.09 (5.33)
Peru	0.67*** (0.13)	5.15***	0.19 (0.87)	26.56***	4.07 (0.27)

Note: Robust standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01
Data source: Bloomberg (Generic Government Rates)

Table 4.9: Regression analysis

Change of 10-year Government Yields for over a year (2017-2018) for BRICS and the U.S. during QT1

	October 2017	**October 2018**	**Change in bps**	**Change in %**
USA	2.29	3.14	84.93	0.85
Brazil	9.99	10.21	21.98	0.22
India	6.86	7.85	99.00	0.99
China	3.88	3.52	-35.50	-0.36
Russia	7.60	8.60	99.50	1.00
South Africa	9.09	9.38	28.50	0.29
Mean	**7.48**	**7.91**	**42.70**	**0.43**

Data source: Bloomberg (Generic Government Rates)

Table 4.10: Change of 10-year Government Yields for over a year (2017-2018) for BRICS and the U.S. during QT1

Change of 10-year Government Yields for over a year (2017-2018) for the Selected EM

	October 2017	October 2018	Change in bps	Change in %
Hungary	2.47	3.64	116.90	1.17
Mexico	7.27	8.88	161.00	1.61
Greece	5.43	4.21	-121.50	-1.22
Turkey	5.28	7.47	219.20	2.19
Peru	5.38	5.89	51.00	0.51
Mean	**5.17**	**6.02**	**85.32**	**0.85**

Data source: Bloomberg (Generic Government Rates)

Table 4.11: Change of 10-year Government Yields for over a year (2017-2018) for the Selected EM

similar occurred in Hungary with a 1.17% annual increase in 10-year government bond yields.

4.6 Conclusion

The empirical approach presented that the U.S. long-term government bond yields have a spillover effect to the Emerging Market economies. This international transmission mechanism came with financial globalization of capital flows and we can observe it in monetary policy decision making as well. We present the spillover effect among chosen countries through the long-end yields co-movement. The result indicates that the strongest correlation was related to the BRICS countries for India (0.90) and Brazil (0.72) during QT1. While, related to the other selected countries, the co-movement was significant for Peru (0.63) and Hungary (0.57) during QT1. We assume the same market bahaviour of yields for QT2 (2022-2025). We also showed that the USA has fulfilled the three preliminary conditions of QT implementation, such as (a.) high GDP% (5.5%), (b.) low unemployment (3.80%), and (c.) policy interest rate hike or over the zero-lower bound (0.50 bp).

Bibliography

[1] Alexander S., Kapovitch V., and Petrunin A. *Alexandrov geometry.* preliminary version no.1, 2019.

[2] Ambrosio L. and Gigli N. *A user's guide to optimal transport.* Modeling and optimisation of flows on networks, 1–155, Lecture Notes in Math., 2062, Springer., 2013.

[3] Ambrosio L., Gigli N., and Savaré G. "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below". In: *Invent. Math. 195, 289–391.* (2014).

[4] Ambrosio L., Gigli N., and Savaré G. *Gradient flows in metric spaces and in the space of probability measures.* Birkhäuser Verlag, 2005.

[5] Ambrosio L., Gigli N., and Savaré G. "Metric measure spaces with Riemannian Ricci curvature bounded from below". In: *Duke Math. J. 163, 1405–1490.* (2014).

[6] Ambrosio L., Savaré G., and Zambotti L. "Existence and stability for Fokker-Planck equations with log-concave reference measure". In: *Probab. Theory Related Fields 145, 517–564.* (2009).

[7] Aragón Artacho F.J., Campoy, R., and Tam, M.K. "The Douglas–Rachford algorithm for convex and nonconvex feasibility problems". In: *Math Meth Oper Res 91, pp. 201–240.* (2020).

[8] Ardia D., Guidotti E., and Kroencke T.A. "Efficient estimation of bid–ask spreads from open, high, low, and close prices". In: *Journal of Financial Economics* 161 (2024), p. 103916. DOI: `https://doi.org/10.1016/j.jfineco.2024.103916`.

[9] Arora J. *Introduction to Optimum Design.* Elsevier, 2017.

[10] Attouch, H. *Variational convergence for functions and operators.* Pitman (Advanced Publishing Program) Applicable Mathematics Series, Boston, MA, 1984.

[11] Azouzi, L. "A diagonal principle for nets". In: *arXiv:2304.04189* (2023).

[12] Bačák M. "Convergence of nonlinear semigroups under nonpositive curvature". In: *J. Appr. Theory* 367 (2015), pp. 3929–3953.

[13] Bačák M. *Convex Analysis and Optimization in Hadamard Spaces*. De Gruyter, 2014. ISBN: 9783110361629.

[14] Bagnell D. Convex Optimization and Subgradients, Carnegie Mellon University. Lecture notes. 2014.

[15] Ball K., Carlen E.A., and Lieb E.H. "Sharp uniform convexity and smoothness inequalities for trace norms". In: *Invent. Math. 115, 463–482.* (1994).

[16] Balser A. and Lytchak A. "Centers of convex subsets of building". In: *Ann. Global Anal. Geom.* 28 (2005), pp. 201–209.

[17] Bauschke H. and Moursi W. *An Introduction to Convexity, Optimization, and Algorithms*. Society for Industrial and Applied Mathematics, 2023.

[18] Berg I. D. and Nikolaev I. G. "A K-quadrilateral cosine characterization of Aleksandrov spaces of curvature bounded above". In: *Arxiv - Metric Geometry (math.MG)* (2015), pp. 1–46.

[19] Bernanke B. *21st Century Monetary Policy: The Federal Reserve from the Great Inflation to COVID-19*. W.W.Norton and Co., 2003.

[20] Bertsekas E. "On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators". In: *Mathematical Programming* 55 (1989), pp. 293–318.

[21] Biais B. et al. "Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market". In: *Review of Economic Studies* 72 (2005), pp. 287–312.

[22] Bidwell E.W. *Vector Analysis: A textbook for the use of students of mathematics and physics*. Yale University Press, 1901.

[23] Bordo M.D., Erceg C., and Michaels R. "Three Great American Disinflations". In: *NBER Working Paper* (2007).

[24] Boyd S. et al. "Performance bounds and suboptimal policies for multi-period investment". In: *Foundations and Trends in Optimization* (2012).

[25] Bredies K., Chenchene E., and Naldi E. "Graph and distributed extensions of the Douglas-Rachford method". In: *SIAM Journal on Optimization 34(2):1569-1594* (2022).

[26] Bredies K. and Sun H. "Preconditioned Douglas-Rachford splitting methods for convex-concave saddle-point problems". In: *SIAM Journal of Numerical Analysis 53(1):421–444.* (2015).

[27] Bridson, M.R. and Haefliger, A. *Metric Spaces of Non-Positive Curvature*. Springer-Verlag, Berlin–Heidelberg, 1999.

[28] Burago D., Burago Y., and Ivanov S. *A course in metric geometry.* Cambridge University Press, 2001.

[29] Cecchetti S. and Tucker P. "Understanding how central banks use their balance sheets: A critical categorization". In: *CERP Policy Research* (2021).

[30] Combettes P.L. "Resolvent and Proximal Compositions". In: *Set-Valued Var. Anal* (1964).

[31] Crandall M. G. and Liggett T. M. "Generation of semi-groups of nonlinear transformations on general Banach spaces". In: *Amer. J. Math.* 93 (1971), pp. 265–298.

[32] Crawley, E. et al. "Substitutability between Balance Sheet Reductions and Policy Rate Hikes: Some Illustrations and a Discussion". In: *FEDS Notes* (2022).

[33] D'Amico, S. and Seida, T. "Unexpected Supply Effects of Quantitative Easing and Tightening". In: *FRB of Chicago Working Paper No. 2020-17* (2020).

[34] Dal Maso, G. *An introduction to* Γ*-convergence.* Birkhäuser Boston Inc., Boston, MA, 1993.

[35] Douglas J. and Rachford H. "On the numerical solution of heat conduction problems in two and three space variables". In: *Transactions of the American mathematical Society* 82 (1956), pp. 421–439.

[36] Eichengreen B. "America's Not-so-Great Inflation". In: *Project Syndicate* (2022).

[37] Eichengreen B. "What is Quantitative Tightening?" In: *Reserve Bank of St. Louis* (2019).

[38] Ekeland I. "Nonconvex minimization problems". In: *Bull. Amer. Math. Soc.* 1 (1979), pp. 443–474.

[39] Erbar M., Kuwada K., and Sturm K.T. "On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces". In: *Preprint (2013). Available at* arXiv:1303.4382 (2014).

[40] Fan J., Qin X., and Tan B. "Convergence of an Inertial Shadow Douglas-Rachford Splitting Algorithm for Monotone Inclusions". In: *Numerical Functional Analysis and Optimization* (2021).

[41] Gagnon J.E. et al. "The Financial Market Effects of the Federal Reserve's Large-Scale Asset Purchases". In: *International Journal of Central Banking 7(1) (March): 3-44* (2011).

[42] Gál H. and Juhász A. "Quantitative Tightening: Theory, Research, and Impact on Selected Emerging Market Economies". In: *Journal of Central Banking Theory and Practice Volume 14 (2025): Issue 1 (January 2025)* (2025).

[43] Gál H. and Lovas A. “Ergodic Behavior of Returns in a Buy Low and Sell High Type Trading Strategy”. In: *In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds) Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2022. Springer, Cham.* (2022).

[44] Gál H. and Pálfia M. “Convergence of semi-convex functions in CAT(1)-spaces”. In: *Calculus of Variations and Partial Differential Equations. 63, pp. 1-20.* (2024).

[45] Gigli N. “On the heat flow on metric measure spaces: existence, uniqueness and stability”. In: *Calc. Var. Partial Differential Equations 39, 101–120.* (2010).

[46] Gigli N. “The splitting theorem in non-smooth context”. In: *Preprint available at* arXiv:1302.5555 (2013).

[47] Gigli N., Kuwada K., and Ohta S. “Heat flow on Alexandrov spaces”. In: *Comm. Pure Appl. Math. 66, 307–331.* (2013).

[48] Giselsson P. and Boyd S. “Diagonal Scaling in Douglas-Rachford Splitting and ADMM”. In: *Conference: 2014 IEEE 53rd Annual Conference on Decision and Control (CDC)* (2014).

[49] Goldstein A. “Convex programming in Hilbert space”. In: *Bull.Amer.Math.Soc* (1964).

[50] Goodhart C. “Changing Role of Central Banks”. In: *FMG Special Papers sp197, Financial Markets Group* (2010).

[51] Gromov M. “Hyperbolic groups”. In: *Essays in group theory. Math. Sci. Res. Inst. Publ. 8. New York: Springer. pp. 75–263.* (1987).

[52] Haan J. de, Schoenmaker D., and Wierts P. *Financial markets and institutional: A European perspective.* Cambridge University Press, 2020.

[53] Hamilton J.D. and Wu J.C. “The Effectiveness of Alternative Monetary Policy Tools in a Zero Lower Bound Environment”. In: *Journal of Money, Credit and Banking 44(S1) (February): pp. 3–46.* (2012).

[54] Ichinose T. et al. “The norm convergence of the Trotter-Kato product formula with error bound”. In: *Comm. Math. Phys.* 221(3):499–510 (2001).

[55] Jordan R., Kinderlehrer D., and Otto F. “The variational formulation of the Fokker–Planck equation”. In: *SIAM J. Math. Anal. 29, 1–17.* (1998).

[56] Jost J. “Convex functionals and generalized harmonic maps into spaces of non-positive curvature”. In: *Comment. Math. Helv. 70, 659–673.* (1995).

[57] Jost J. “Equilibrium maps between metric spaces”. In: *Calc. Var. PDE* 2 (1994), pp. 173–204.

[58] Jost J. "Nonlinear Dirichlet forms. Nonlinear Dirichlet forms". In: *Amer. Math. Soc. Providence, RI* (1998), pp. 1–47.

[59] Joyce, M.S. et al. "The Financial Market Impact of Quantitative Easing in the United Kingdom". In: *International Journal of Central Banking 7(3) (Sept.):113-162* (2011).

[60] Kato T. "On the Trotter-Lie product formula". In: *Proc. Japan Acad.* (1974).

[61] Kato T. "Trotter's product formula for an arbitrary pair of self-adjoint contraction semigroups". In: *In Topics in functional analysis (essays dedicated to M. G. Kreĭn on the occasion of his 70th birthday), volume 3 of Adv. in Math. Suppl. Stud., pages 185–195. Academic Press, New York-London* (1978).

[62] Kell M. "Analysis and Geometry of Metric Spaces". In: *Uniformly convex metric spaces* 2 (2014), pp. 359–380.

[63] Krishnamurthy, A. and Vissing-Jørgensen, A. "The Ins and Outs of LSAPs". In: *In Economic Symposium Conference Proceedings. Jackson Hole, WY: Federal Reserve Bank of Kansas City* (2013).

[64] Kuwae K. "Resolvent Flows for Convex Functionals and p-Harmonic Maps". In: *Analysis and Geometry in Metric Spaces* 3 (2015), pp. 46–72.

[65] Kuwae K. "Variational convergence over p-uniformly convex spaces, in Geometric Aspect of Partial Differential Equations and Conservation Laws". In: *RIMS Kokyuroku* 1842 (2013), pp. 22–39.

[66] Kuwae K. and Shioya T. "Variational convergence over metric spaces". In: *Trans. Amer. Math. Soc.* 360 (2008), pp. 35–75.

[67] Kyle D. Allen and Hein E. Scott. "Unusual Changes in the U.S. Treasury Security Market During the Fourth Round of Quantitative Easing". In: *Journal of Central Banking Theory and Practice, Vol. 12(3), pages 5-22.* (2023).

[68] Hopper L. J. "How the Fed Is Reducing Its Balance Sheet—and Why". In: *Federal Reserve Bank of St. Louis* (2018).

[69] Lee, J.M. *Connections. In: Introduction to Riemannian Manifolds.* Springer, 2018.

[70] Lee, J.M. *Introduction to Topological Manifolds*. Springer, 2011.

[71] Lim J., Mohapatra S., and Stocker M. "Tinker, Taper, QE, Bye? The Effect of Quantitative Easing on Financial Flows to Developing Countries". In: *Policy Research Working Paper; No. 6820. World Bank, Washington, DC. World Bank* (2014).

[72] Lions P.L. and Mercier B. "Splitting algorithms for the sum of two nonlinear operators". In: *SIAM Journal on Numerical Analysis* 16 (1979), pp. 964–979.

[73] Chang Liu. "Gradient Flow.Tsinghua University". 2017.

[74] Löfgren K.G., Persson T., and Weibull J.W. "Markets with Asymmetric Information: The Contributions of George Akerlof, Michael Spence and Joseph Stiglitz". In: *The Scandinavian Journal of Economics, Vol. 104, No. 2 (Jun., 2002), pp. 195-211* (2002).

[75] Luenberger D. *Optimization by Vector Space Methods.* John Wiley & Sons, Inc., 1969.

[76] Mahey P. and Lenoir A. "A survey on operator splitting and decomposition of convex programs". In: *RAIRO - Operations Research* (2017).

[77] Mayer U. F. "Gradient flows on nonpositively curved metric spaces and harmonic maps". In: *Comm. Anal. Geom.* 6 (1998), pp. 199–253.

[78] Milgrom P. and Stokey N. "Information, Trade, and Common Knowledge". In: *Journal of Economic Theory* 26 (1982), pp. 17–27.

[79] Fadili J. Molinari C. Liang J. "Convergence Rates of Forward Douglas–Rachford Splitting Method". In: *Journal of Optimization Theory and Applications* (2019).

[80] Moore E. Ramon and Cloud J. Michael. *Types of convergence in function spaces. In: Computational Functional Analysis.* Woodhead Publishing, 2007.

[81] Mosco U. "Composite media and asymptotic Dirichlet forms". In: *J. Funct. Anal.* 123 (1994), pp. 368–421.

[82] Mulaahmetović I. "Quantitative Easing and Macroeconomic Performance in the United States". In: *Journal of Central Banking Theory and Practice, Vol. 11(3), pages 79-98.* (2022).

[83] Muratori M. and Savaré G. "Gradient flows and Evolution Variational Inequalities in metric spaces. I: Structural properties". In: *J. Func. Anal.* (2020).

[84] Neidhardt H. and Zagrebnov V. A. "On error estimates for the Trotter-Kato product formula". In: *Lett. Math. Phys.* 44(3):169–186 (1998).

[85] Neidhardt H. and Zagrebnov V. A. "Trotter-Kato product formula and operator-norm convergence". In: *Communications in Mathematical Physics* (1999).

[86] O'Donoghue B., Stathopoulos G., and Boyd S. "A splitting method for optimal control". In: *IEEE Transactions on Control Systems Technology* (2012).

[87] Ohta S.-I. "Convexities of metric spaces". In: *Geom. Ded.* (2007), pp. 225–250.

[88] Ohta S.-I. "Uniform convexity and smoothness, and their applications in Finsler geometry". In: *Math. Ann. 343 (2009), 669–699.* (2009).

[89] Ohta S.-I. and Pálfia M. "Discrete-time gradient flows and law of large numbers in Alexandrov spaces". In: *Calc. Var. PDE* 54 (2015), pp. 1591–1610.

[90] Ohta S.-I. and Pálfia M. “Gradient flows and a Trotter-Kato formula of semi-convex functions on CAT(1)-spaces”. In: *Amer. J. Math.* 139 (2017), pp. 937–965.

[91] Ohta S.I. and Sturm K.T. “Non-contraction of heat flow on Minkowski spaces”. In: *Arch. Ration. Mech. Anal. 204, 917–944.* (2012).

[92] Otto F. “The geometry of dissipative evolution equations: the porous medium equation”. In: *Partial Differential Equations 26, 101–174.* (2001).

[93] Parikh N. and Boyd S. *Proximal Algorithms*. Foundations and Trends in Optimization, 2013.

[94] G Peña. “Raising Interest Rates for Improving Income”. In: *Journal of Central Banking Theory and Practice, Vol. 12(3), pages 199-217.* (2023).

[95] Peyré G. *Douglas Rachford Proximal Splitting*. 2014. URL: `https://www.numerical-tours.com/matlab/optim_4b_dr/`.

[96] Rogava Dzh. “Error bounds for Trotter-type formulas for self-adjoint operators”. In: *Funct. Anal. Appl.* (1993).

[97] Ryu K. E. and Boyd S. “Primer on monotone operator methods”. In: *Appl. Comput. Math, 15(1):3-43* (2016).

[98] Sarhan M. Musa. *Fundamentals of Technical Mathematics*. Elsevier, 2015.

[99] Savaré G. “Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds”. In: *C. R. Math. Acad. Sci. Paris 345, 151–154* (2007).

[100] Svaiter B. “A simplified proof of weak convergence in Douglas–Rachford method”. In: *Operations Research Letters* (2019).

[101] Toyoda T. “A non-geodesic analogue of Reshetnyak’s majorization theorem”. In: *Analysis and Geometry in Metric Spaces* (2023).

[102] Trotter H. F. “On the product of semi-groups of operators”. In: *Proc. Amer. Math. Soc.* (1959).

[103] Villani C. *Optimal transport, old and new*. Springer-Verlag, Berlin, 2009.

[104] Villani C. “Topics in optimal transportation”. In: *American Mathematical Society, Providence, RI.* (2003).

[105] Wei B. “How Many Rate Hikes Does Quantitative Tightening Equal?” In: *Federal Reserve of Atlanta’s Policy Hub. Center for Quantitative Economic Research, No. 11–2022, July 2022.* (2022).

[106] Yokota T. “Convex functions and barycenter on CAT(1)-spaces of small radii”. In: *J. Math. Soc. Japan* 68 (2016), pp. 1297–1323.

[107] Zagrebnov V. A. “Trotter product formula on Hilbert and Banach spaces for operator-norm convergence”. In: *Universitätsverlag Potsdam Proceedings of the XI international conference Stochastic and Analytic Methods in Mathematical Physics Proceedings of the XI international confer-ence stochastic and analytic methods in mathematical physics* (2020).

[108] Zill G. Dennis and Cullen R. Michael. *Advanced engineering mathematics (3rd ed.)* Jones & Bartlett Learning, 2006.