UNIVERSITÉ DE LIMOGES

ÉCOLE DOCTORALE S2IM n°521

FACULTÉ DES SCIENCES ET TECHNIQUES

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE LIMOGES

Discipline : Mathématiques et Applications

présentée et soutenue par

Van Vu NGUYEN

le 30 Septembre 2016

Méthode de Newton Revisitée pour les Equations Généralisées

Thèse dirigée par: Samir ADLY et HUYNH Van Ngai

JURY :

Samir ADLY	Professeur, Université de Limoges	Directeur de thèse
Paul ARMAND	Professeur, Université de Limoges	Examinateur
Hedy ATTOUCH	Professeur émérite, Université de Montpellier	Rapporteur
Radu Ioan BOT	Assoc. Prof. Dr. habil., Université de Vienne	Examinateur
Michel H. GEOFFROY	Professeur, Université des Antilles	Rapporteur
Van Ngai HUYNH	Assoc. Prof. Dr. habil., Université de Quy Nhon	Codirecteur de thèse
Abderrahim JOURANI	Professeur, Université de Bourgogne	Examinateur
Michel THERA	Professeur émérite, Université de Limoges	Président

To my parents,

Acknowledgements

The author would like to express his deep gratitude to the supervisors, Prof. Samir Adly and Assoc. Prof. Dr. habil. Huynh Van Ngai, who had instructed him during the period of his thesis. Without their unconditional enthusiasm guidances, this dissertation would never have been able to complete.

The author is very grateful the jury, who had spent their precious time to read this dissertation, for their insightful comments and advices. I thank the referees, Prof. Hedy Attouch and Prof. Michel H. Geoffroy, who agreed to write a review of my thesis. The heartfelt thanks are due to Prof. Paul Armand, Assoc. Prof. Dr. habil. Radu I. Boţ, Prof. Abderrahim Jourani and Prof. Michel Thera for the acceptance to be a member of the jury.

This thesis has been completed at the Xlim laboratory, Faculty of Sciences and Technologies in the University of Limoges. I greatly appreciate the librarian, the secretariats and other people here for their kind helps to the work as well as the life in a professional scientific environment. I thank the colleagues and friends at the Xlim Laboratory (especially, I want to mention Isaï Lankoande and Florent Nacry) for their friendly companionship and cooperation.

My sincere acknowledgements are also dedicated to the University of Limoges and the Region Limousin, that have given the financial supports to finish the doctoral thesis.

To teachers, colleagues and friends who have been working in the University of Quy Nhon, the author is grateful them for their encouragements as well as their kind helps of solving his own trouble.

Far from a familiar environment, I am pleasant to have many friends together with whom I have fantastic experiences. Particularly, I would like to thank my gracious friends: Duc, Thanh, Nguyen, Duy, Bang, Thuy, Tu, Nga, Ha, Quan, Thieu, etc.

Finally, the author is unlimited grateful to his family. I am indebted to my parents, who carried me to this world and have brought me up today. Thanks my brothers and my sister, who have been besides and emboldened me through my life.

Contents

List of Figures

Chapter 1

General Introduction

Introduction française

Résoudre des équations non-linéaires est un problème fondamental en mathématiques qui a une longue histoire dans la littérature. Son importance est due au fait que les équations non-linéaires apparaissent dans de nombreux domaines d'applications, non seulement en mathématiques appliquées mais aussi en physique, sciences de l'ingénieur, biomathématique. De nombreuses méthodes numériques ont été étudiées pour l'approximation de solutions de tels problèmes. Il est bien connu dans la littérature que la méthode de Newton (ou Newton-Raphson ou Newton-Raphson-Simpson) classique ainsi que ses extensions sont parmi les plus efficaces. Ce succès est lié notamment à la vitesse de convergence de la suite itérative générée pour un choix approprié de point de départ. Pour plus de détails, le lecteur pourra consulter les livres suivants [15, 18, 52, 54, 65].

Un système d'équations non-linéaires est un système de la forme $f(x) = 0$ où $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ est une fonction lisse donnée. Localement, la théorie classique de la méthode de Newton indique que si $f(x^*) = 0$, si le gradient $Df(x^*)$ de f au point x^* est *inversible* et si Df est *lipschitzienne* autour de x^*, alors il existe un voisinage Ω^* de x^* ayant la propriété suivante : pour chaque point de départ $x_0 \in \Omega^*$, l'algorithme est défini par

$$x_{k+1} = x_k - [Df(x_k)]^{-1} f(x_k) \tag{1.1}$$

converge *Q-quadratiquement* vers x^*. En général, lorsque Df n'est pas lipschitzienne, le taux de convergence est simplement *sur-linéaire* (cf. [46]).

La convergence locale exige des informations autour de la solution x^*, ce qui est généralement inconnu. Il est donc important d'étudier d'autres types de résultats de convergence avec des hypothèses ne nécessitant pas d'informations locales autour de x^*. L'un des plus célèbres résultat dans ce sens est le *théorème de Kantorovich* (voir, par exemple, [14, 51, 64]). Selon l'article [64], il semblerait que L.V. Kantorovich ait donné deux preuves de ce théorème. Il a d'abord utilisé les relations de récurrence pour le prouver, puis ensuite reformulé par la technique de *fonction majorante*. Ce théorème fournit des conditions suffisantes imposées sur les données initiales seulement[1] assurant à la fois l'existence de la suite de Newton ainsi que des bornes d'erreur pour la convergence (généralement nommé *R-quadratique*). Plus tard dans l'article [39], les auteurs ont établi ce résultat avec *borne d'erreur optimale* en utilisant l'approche de Kantorovich. Nous renvoyons à Galántai [38] pour plus de détails sur le théorème de Newton-Kantorovich. Il existe de nombreuses applications et extensions de ce théorème dans différentes situations (voir, par exemple, [32, 35, 36, 84, 88]).

En ce qui concerne la méthode de Newton classique, le régularité de f autour de la solution x^* est essentielle. Dans un contexte plus large, il existe des généralisations pour la résolution d'équations impliquant des fonctions non-lisses telles que la méthode semismooth de Newton ([53, 66, 81]). De plus, ces résultats peuvent être généralisés pour être appliqués à un modèle plus général appelé *équations généralisées*[2]. Historiquement, ces études remontent aux travaux de Bakushinskii [10], Josephy [50], et Robinson [70–72]. Le lecteur est renvoyé au monographe de Izmailov et Solodov [46] ainsi qu'à l'article récent [47] pour plus de détails où un survol des derniers résultats est donné au lien avec l'optimisation et l'analyse variationelle.

Dans une forme abstraite, une *équation généralisée* est définie comme suit

$$\text{trouver} \quad x \in X \quad \text{tel que} \quad 0 \in f(x) + F(x). \tag{1.2}$$

Ici, $f : X \longrightarrow Y$ est une fonction continue entre deux espaces de Banach X, Y tandis que $F : X \rightrightarrows Y$ est une multi-application. Tout au long de cette thèse, f est supposée être de classe C^1 au moins sur un sous-ensemble convexe ouvert de X et F est supposée avoir un graphe fermé. Le modèle (1.2) couvre un grand nombre d'applications en mathématiques et ingénierie. Selon la forme spécifique de la multi-application F, le problème (1.2) devient un problème d'admissibilité ($F(x) \equiv K$), une inéquation variationnelle ou un problème de complémentarité (le cas où $F(x)$ coïncide avec le cône normal à un ensemble fermé convexe ou un cône convexe fermé, respectivement). D'autres discussions sur ces sujets peuvent être

[1] il ne comporte pas d'informations sur la solution elle-même
[2] cette terminologie est due à S.M. Robinson

trouvées dans la littérature, par exemple [30, 31, 53].

Dans certaines situations, il est possible de transformer le problème (1.2) en une équation, en utilisant par exemple les fonctions normales introduites par S.M. Robinson [73] pour résoudre les inéquations variationnelles. Par exemple, le solveur *PATH*, l'un des solveurs le plus populaire pour résoudre numériquement les problèmes de complémentarité mixtes [37], est basé sur les fonctions normales et la méthode de Newton non-lisse.
Malheureusement, cette technique ne peut pas être adaptée pour l'équation généralisée de la forme (1.2) avec une application multivoque arbitraire $F(\cdot)$. Il serait peut-être naturel de traiter le problème (1.2) directement au lieu de le transformer en une équation. En particulier, une méthode de type Newton pour traiter (1.2) dans le cas des inéquations variationnelles (VIs) a été proposé dans [10] (sous l'hypothèse de monotonie de f) et [50]. Dans l'approche des articles [10, 50], nous constatons que le cône normal pourrait être remplacé par un opérateur quelconque F (à graphe fermé).

L'algorithme de type Newton appliqué à une forme abstraite (1.2), connu sous le nom *méthode de Josephy-Newton* (JNM), est paru dans les travaux de A.L. Dontchev [22, 23]. Plus précisément, l'algorithme commence en un point x_0 à proximité d'une solution et génère une suite d'approximation (x_k) en résolvant le sous-problème

$$0 \in f(x_k) + Df(x_k)(x - x_k) + F(x) \tag{1.3}$$

pour obtenir une nouvelle itération x_{k+1}. L'approximation linéaire $f(x_k) + Df(x_k)(\cdot - x_k)$ de f peut être traitée dans un cadre plus général et abstrait

$$0 \in A_k(x_{k+1}, x_k) + F(x_{k+1}), k = 0, 1, \ldots \tag{1.4}$$

où $A_k : X \times X \rightrightarrows Y$ est une suite de multi-applications, satisfaisant certaines hypothèses générales pour approcher f dans un voisinage de la solution. Le cas particulier où A_k dans (1.4) est un opérateur univoque a été discuté dans le monographe [24], alors que le cas général d'approximation multivoque a été étudié dans l'article Adly et al. [5]. Pour plus d'informations sur les méthodes de type Newton (exacte et inexacte) pour les équations généralisées, nous renvoyons le lecteur aux références suivantes [13, 19, 25, 48, 80, 85, 86].

Dans [21], A.L. Dontchev a donné des résultats (locales et semi-locales) impliquant une version étendue du théorème de Kantorovich dans le cadre de (1.2). Les preuves étaient basées sur la notion de régularité métrique locale pour les applications multivoques, qui joue un rôle similaire à l'inversibilité de la dérivée du premier ordre dans la version originale donnée par Kantorovich. Une approche similaire se trouve dans [67]. Cependant, l'approche de Dontchev, dans son théorème semi-local, nécessite les informations non seulement sur le point de départ x_0 mais aussi sur x_1. Alors que dans [67], les auteurs se sont concentrés sur le comportement de $f + F$ autour du point de référence $(\bar{x}, \bar{y}) \in \mathrm{Gr}(f + F)$. Nous discuterons dans le Chapitre 3 de notre théorème de type Kantorovich pour résoudre (1.2) ainsi que d'une comparaison avec les approches présentées dans [21] et [67]. Pour le cas local, nous montrons que sous les hypothèses relatives à la régularité métrique de $Df(x^*)(\cdot) + F(\cdot)$ (x^* est supposé être une solution de (1.2)) ainsi que des informations sur la dérivée seconde $D^2 f$ dans un voisinage de x^*, il existe une suite (x_k) générée par (1.3) qui converge *Q-quadratiquement* vers x^*. D'autre part, en invoquant la régularité métrique de $Df(x_0)(\cdot) + F(\cdot)$ (ici x_0 indique le point de départ) et le comportement de $D^2 f$ autour x_0, l'algorithme (1.3) induit une suite qui converge *R-quadratiquement* vers x^*.

Outre le théorème de Kantorovich, les théories de (α, γ)-Smale [12, 77] représentent également des résultats fondamentaux en analyse numérique. Appliquée à une équation $f(x) = 0$, où f est supposée être une fonction analytique, les (α, γ)-théories de Smale fournissent des critères suffisants garantissant la convergence quadratique (vers une solution) pour la suite de Newton avec une estimation seulement en la solution (pour la γ-théorie) et au point de départ (pour la α-théorie). A titre de comparaison, le théorème de Kantorovich utilise les informations des dérivées du premier et second ordre Df et $D^2 f$ dans un voisinage de x_0, tandis que celle de α-Smale exige l'analyticité de f en x_0 et utilise toutes les dérivées d'ordre supérieurs $D^j f(x_0)$, $j \in \mathbb{N}$. Dans certaines situations, la seconde approche est plus commode en pratique que la première car le maximum de la norme $\|D^2 f(\cdot)\|$ sur un voisinage du point de départ x_0 ne peut-être calculé facilement.

Après le travail fondateur de S. Smale, de nombreux chercheurs ont essayé d'améliorer et d'étendre les (α, γ)-théories classiques pour une grande classe de problèmes, voir par exemple [17, 76, 83, 87]. Néanmoins, il n'existe pas de résultats, à notre connaissance, qui étudient le problème (1.2) lié à l'approche de Smale. Ce sera l'objet du Chapitre 3, qui est basé sur le papier [6], considéré comme le premier à adapter les théories de Smale pour les équations généralisées (1.2).

Comme mentionné ci-dessus, la méthode de Josephy-Newton utilise une certaine *linéarisation partielle* de la somme $f + F$ à chaque itération. Plus précisément, lorsque x_k est connu, on remplace f par sa linéarisation en ce point et on considére le problème auxiliaire (1.3) au lieu de (1.2). Ceci est également le cas pour d'autres méthodes appliquées aux équations généralisées (1.2), comme la méthode de Newton inexacte [25] et les méthodes de type quasi-Newton [9]. Dans la plupart de ces travaux, la linéarisation partielle n'est effectuée que sur la partie univoque f. Dans leur article [39], M. Gaydu et M.H. Geoffroy ont proposé une méthode numérique locale pour laquelle à la fois f et F pourraient être approximées. Ceci a été rendu possible en utilisant le concept de *différenciation généralisée* introduite par C.H.J. Pang [49]. L'algorithme produit une suite (x_k) en résolvant successivement les sous-problèmes de la forme

$$0 \in f(x_k) + Df(x_k)(x_{k+1} - x_k) + H(x_{k+1} - x_k) + F(x_k), k = 0, 1, \ldots \qquad (1.5)$$

où $H : X \rightrightarrows Y$ est une *multi-application positivement homogène*, qui est une dérivée stricte de F [39, 49] à la solution $x^* \in X$ de (1.2). La stratégie clé de (1.5) est de prendre $f(x_k) + Df(x_k)(\cdot - x_k) + H(\cdot - x_k) + F(x_k)$ comme une approximation pour $f + F$ tout au long du processus itératif. Notons que dans l'algorithme (1.5), $H(\cdot)$, dérivée de la multi-application de F en x^*, ne dépend pas de k (constante tout au long des itérations). Néanmoins, une telle solution est couramment inconnue dans la pratique.

En ce qui concerne nos travaux dans cette thèse, nous proposerons au Chapitre 4 une extension de (1.5) pour résoudre l'équation généralisée (1.2). Soit $\mathcal{H} : X \longrightarrow \mathscr{PH}(X, Y)$ une application donnée de X vers l'ensemble $\mathscr{PH}(X, Y)$ de toutes les applications homogènes entre X et Y. A partir de x_k, on met à jour le prochain itéré x_{k+1} en résolvant le problème auxiliaire

$$0 \in f(x_k) + Df(x_k)(x - x_k) + \mathcal{H}(x_k)(x - x_k) + F(x_k). \qquad (1.6)$$

Il est clair que (1.5) peut être regardé comme un cas particulier de (1.6) en posant $\mathcal{H}(x) \equiv H$. Observons que le terme d'approximation $\mathcal{H}(x_k)(x - x_k) + F(x_k)$ dans (1.6) varie à chaque itération. En supposant certaines hypothèses sur l'uniformité de la "composante homogène" de $\mathcal{H}(\cdot)$ ainsi que la propriété de régularité métrique de $Df(x_0)(\cdot - x_0) + \mathcal{H}(x_0)(\cdot - x_0)$, nous prouvons que le schéma numérique (1.6) génère au moins une suite convergente *R-linéairement* vers une solution (pour la *convergence semi-locale*, voir le théorème 4.6). La même conclusion est également valide dans l'*analyse locale*, lorsque une solution x^* de (1.2) est considérée au lieu de x_0. Un raffinement de la convergence linéaire locale est obtenue si une hypothèse plus forte sur la différentiabilité en x^* est vérifiée.

Au cours des dernières décennies, l'idée d'étudier des méthodes itératives sur les *variétés riemanniennes* a été développée par de nombreux auteurs. Ces recherches ont été motivées par de nombreux problèmes apparaissant dans plusieurs applications, telles que l'optimisation avec contraintes, la décomposition singulière, ou l'approximation matricielle. Par exemple, le problème de valeurs propres peut être reformulé sous la forme de minimisation d'une fonction à valeurs réelles lisse définie sur une variété appropriée (le *quotient de Rayleigh* sur la *sphère unité* [1]). En ce qui concerne les motivations et les applications des méthodes sur les variétés, voir les livres [1, 79] et aussi les articles suivants [2, 3, 27, 43, 58, 68, 78].

Jusqu'à présent, il existe un grand nombre de résultats concernant les méthodes de type Newton ainsi que leurs extensions sur des variétés lisses. Par exemple, les articles [8, 33, 34] étendent le théorème de Kantorovich et [16] affine la théorie de Smale afin de trouver une singularité d'un champ de vecteurs lisse. Si f est une fonction à valeurs réelles lisse définie sur une variété $\mathcal{M}$, alors tout point critique de f est une singularité de son gradient $\operatorname{grad} f$, ce qui résout l'équation $\operatorname{grad} f(p) = 0$. En outre, l'article [82] considére l'inclusion $0 \in f(p) + C$, où $C \subset \mathbb{R}^n$ est un cône et $f : \mathcal{M} \longrightarrow \mathbb{R}^n$ est au moins de classe C^2. En utilisant la méthode introduite dans [69] et la technique dans [83], les auteurs de [82] ont étudié les théorèmes de Kantorovich et Smale pour la suite d'approximation d'une solution du problème $0 \in f(p) + C$.
Avec les motivations mentionnées ci-dessus, nous traitons le problème de résolution numérique d'une inclusion de la forme

$$0 \in \varphi(p) + \Phi(p), \tag{1.7}$$

où $\varphi : \mathcal{M} \longrightarrow \mathbb{R}^n$ est une application lisse et $\Phi : \mathcal{M} \rightrightarrows \mathbb{R}^n$ est une multi-application ayant un graphe fermé. Remarquons que, (1.7) est réduite à celle étudiée dans [82] en prenant $\Phi(p) \equiv C$. En posant dans (1.7), $\Phi(p) = \{0\}$ où $\varphi = (V_1, \dots, V_m)$, avec $V_1, \dots, V_m$ les composantes de la représentation du champ de vecteurs V par rapport à un certain cadre bien choisi $\{E_1, \dots, E_m\}$ [3], le modèle considéré (1.7) recouvre à nouveau le problème de singularité pour le champ de vecteur dans la littérature.

Dans le même esprit de (1.3), nous allons faire usage de linéarisation partielle afin de générer les approximations successives d'une solution à (1.7). Plus précisément, en ayant une itération p_k, nous allons suivre une rétractation R_k

[3] dans la théorie de la géométrie différentielle, un tel cadre est naturel

à l'étape en cours, et de remplacer (1.7) par l'inclusion suivante

$$0 \in \varphi(p_k) + \mathscr{D}\varphi(p_k)(u) + (\Phi R_k)(u), \tag{1.8}$$

pour obtenir une direction de recherche u_k. Ensuite, nous mettons à jour $p_{k+1} = R_k(u_k)$ comme la nouvelle itération. Grâce à un choix approprié de la partie multivaluée Φ et les rétractions R_k, (1.8) peut être considérée comme une continuation de la méthode de Newton présente dans [16, 34, 82]. En suivant la même stratégie, nous montrons dans le Chapitre 5 la convergence locale et semi-locale de (1.8) avec quelques hypothèses imposées sur la structure de la variété ainsi que le comportement des rétractions R_k. Encore une fois, la notion de la régularité métrique pour les applications multivoques joue un rôle important dans notre analyse.

L'ensemble du contenu de cette thèse est organisé comme suit. Le Chapitre 1 est une introduction à la problèmatique. Le Chapitre 2 rappelle quelques résultats de base ainsi que les notations qui seront utilisées tout au long de notre travail. Le Chapitre 3 est consacré aux théorèmes de convergence de Kantorovich et de Smale pour les équations généralisées. Ceci a fait l'objet d'une publication dans la revue internationale "Journal of Mathematical Analysis and Applications" [6], en collaboration avec S. Adly et H.V. Ngai. Le Chapitre 4 introduit une méthode de type Newton de résolution des équations généralisées (1.2) en approximant à la fois la partie univoque et multivoque. Des résultats concernant la convergence de l'algorithme proposé ont été prouvés. (Ce chapitre est basé sur le manuscrit [7] soumis à la revue internationale "Set-Valued and Variational Analysis"). Le Chapitre 5 prend en compte un algorithme de Newton-type pour la résolution des inclusions qui impliquent des applications multi-valuées définies sur des variétés riemanniennes. Les ingrédients principaux de ce chapitre sont basés sur la géométrie riemannienne, ainsi que des outils de l'analyse variationnelle, où la propriété de régularité métrique est un point clé. (Ce chapitre est basé sur le manuscrit [4] soumis à la revue internationale "Journal of Convex Analysis").

English Introduction

Solving nonlinear equations is a basic problem which has a long history in the literature. Its importance is due to the fact that, nonlinear equations appear in many fields of applications, not only mathematics itself. There were many techniques seeking the solutions of such a problem. It is well-known in the literature that the classical Newton's method and its extensions are among of the most popular and efficient ones. This success is related to the good behavior of convergence of the Newton iterative sequence under a good choice of suitable starting point. We refer to the textbooks e.g. [15, 18, 52, 54, 65] for more details.

Because the phrase "Newton's method" frequently appears in this text, it is advantage for us to abbreviate it as "NM". Particularly, let's consider the equation $f(x) = 0$ where $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is a given smooth map. Locally, the classical theory for NM states that, if $f(x^*) = 0$, the derivative $Df(x^*)$ of f at x^* is *invertible* (resp. the Jacobian $\text{Jac}_f(x^*)$ is *nonsingular*) and Df is *Lipschitz continuous* around x^*, then there exists a neighborhood Ω^* of x^* having the following property: for each starting point $x_0 \in \Omega^*$, the algorithm defined by

$$x_{k+1} = x_k - [Df(x_k)]^{-1} f(x_k) \tag{1.9}$$

converges *Q-quadratically* to x^*. Otherwise, without the Lipschitz continuity property on Df, the rate of convergence might be just *superlinear* (cf. [46]).

The local convergence requires the informations around the solution x^*, which is usually unknown. So, it is important to study other type of convergent results with different assumptions. One of the most famous ones is the classical *Kantorovich's theorem* (see e.g. in [14, 51, 64]). Following the paper [64], it seems that L.V. Kantorovich had given two proofs for this theorem. Firstly, he used recurrence relations to prove it, and then reformulated it by the technique of *majorizing function*. The reference theorem provides some sufficient conditions imposed on the initial data only[4], which ensure both the existence of the Newton sequence and the error bounds for the convergence (usually named as *R-quadratic*). Later in the paper [40], the authors reestablished that result with an *optimal error bound* using the approach of Kantorovich. A survey about Newton-Kantorovich theorem can be found in the paper by Galántai [38]. There are many successful applications and extensions of such a theorem in different situations, for instance, see [32, 35, 36, 84, 88].

Concerning the classical NM, the smoothness of f around a solution x^* is

[4] it does not involve any information about solution itself

essential. In more wider context, there were some generalizations for handling with equations involving nonsmooth maps, such as, the semismooth Newton method (see e.g. Klatte and Kummer [53], Qi and Sun [66], Ulbrich [81]). Furthermore, these frameworks can be generalized to be applied to an extensive model named as *generalized equations*[5] (GE). Historically, such studies trace back to the works of Bakushinskii [10], Josephy [50], and Robinson [70–72]. The readers are referred to the monograph by Izmailov and Solodov [46] and also the recent paper [47] for an overview.
In general form, a GE is defined as follows

$$\text{find} \quad x \in X \quad \text{such that} \quad 0 \in f(x) + F(x). \tag{1.10}$$

Here, $f : X \longrightarrow Y$ is a continuous map between two Banach spaces X, Y while $F : X \rightrightarrows Y$ is a set-valued map (or multifunction). Throughout this thesis, f is always assumed to be C^1 at least on some open convex set of X and F has a closed graph. The model (1.10) covers a lot of applications in mathematics, engineering and sciences. Depending specific form on the set-valued term F, problem (1.10) becomes a feasibility problem ($F(x) \equiv K$), a variational inequality or a complementarity problem ($F(x)$ coincides with the normal cone to a closed convex set or closed convex cone, respectively). Further discussions on those subjects can be found in the literature, e.g. [30, 31, 53].
In some situations, it is possible to transform problem (1.10) into an equation, by using for example the normal maps introduced by S.M. Robinson [73] for solving variational inequalities. For instance, the *PATH* solver, based on normal maps and nonsmooth Newton method, is one of the most popular solver for solving numerically mixed complementarity problems [37].
Unfortunately, this technique can not be adapted for GE of the form (1.10) with an arbitrary set-valued $F(\cdot)$. As an alternative solution, it may be natural to deal directly with (1.10) instead of transforming it to an equation. In particular, a Newton-type method for solving (1.10) in the case of variational inequalities (VIs) was proposed in [10] (under the monotonicity assumption of f) and [50]. We notice that in the approach of the papers [10, 50], the normal cone could be replaced by any set-valued F.

Newton-type algorithm applies to an abstract GE (1.10), known as the *Josephy-Newton method* (JNm), appeared in the works by Dontchev [22, 23]. Precisely, the algorithm starts at a point x_0 nearby a solution, and generates an approximating

[5] this terminology is due to S.M. Robinson

sequence (x_k) by solving the subproblem

$$0 \in f(x_k) + Df(x_k)(x - x_k) + F(x) \tag{1.11}$$

to obtain new iteration x_{k+1}. The linear approximation $f(x_k) + Df(x_k)(\cdot - x_k)$ for f can be treated in a more general abstract setting

$$0 \in A_k(x_{k+1}, x_k) + F(x_{k+1}), k = 0, 1, \ldots \tag{1.12}$$

where $A_k : X \times X \rightrightarrows Y$ is a sequence, which should satisfy some general assumptions to approximate f in a neighborhood of the solution. The special case where A_k in (1.12) is single-valued has been discussed in the monograph [24], while a general consideration for set-valued maps approximation was studied in the paper Adly et al. [5]. Much more on Newton-type schemes (exact and inexact) for GEs can be found in [13, 19, 25, 48, 80, 85, 86].

In [21], A.L. Dontchev has stated some results (both local and semi-local versions) involving extended Kantorovich theorem for the framework of (1.11). His proofs were based on the concept of local metric regularity property for set-valued maps, which played a similar role as the invertibility of the first-order derivative in the original version by Kantorovich. A similar approach is able to be found in [67]. However, Dontchev required in his semi-local theorem the informations not only about the starting point x_0 but also x_1. While in [67], the authors focused on the behavior of $f + F$ around the reference point $(\bar{x}, \bar{y}) \in \mathrm{Gr}(f + F)$. We shall discuss in Chapter 3 our Kantorovich-type theorems for solving (1.10) together with a brief comparison with the ones presented in [21] and [67]. We show that, for local case, under the assumptions related to metric regularity of $Df(x^*)(\cdot) + F(\cdot)$ (x^* is assumed to be a solution of (1.10)) as well as the information on the second derivative $D^2 f$ in a neighborhood of x^*, there exists a sequence (x_k) by (1.11) which converges *Q-quadratically* to x^*. On the other hand, by invoking the metric regularity of $Df(x_0)(\cdot) + F(\cdot)$ (here x_0 indicates the starting point) and the behavior of $D^2 f$ around x_0, the algorithm (1.11) induces a sequence converging *R-quadratically* to a solution.

Beside the theorem of Kantorovich, the so-called (α, γ)-Smale's theories [12, 77] are also fundamental results in numerical analysis. Applied to an equation $f(x) = 0$, where f is supposed to be analytic, the (α, γ)-Smale's theories provided some sufficient criteria guaranteeing the quadratic convergence (to a solution) for the Newton sequence under estimation only at the solution (for γ-theory)

and respectively the starting point (α-theory). As a comparison, Kantorovich's theorem used the information of the first and second order derivatives Df and D^2f in a neighborhood of x_0, while the Smale's α-theorem requires the analyticity of f at x_0 and used all derivatives $D^j f(x_0)$, $j \in \mathbb{N}$. In some situations, the second approach is more convenient in practice than the first one since the maximum of the norm $\|D^2f(\cdot)\|$ over a neighborhood of starting point x_0 could not be easy to compute.
After Smale's work, many researches have tried to improve and extend the classical (α, γ)-theorems into a large class of problems, see e.g. [17, 76, 83, 87]. Nevertheless, there are almost very few results, to the best of our knowledge, that study problem (1.10) related to Smale's approach. This will be the purpose of Chapter 3, which is based on the paper [6], considered to be the first one to adapt Smale's theories for GE (1.10).

Following the aforementioned discussion, the Josephy-Newton framework used some representation of a *partial linearization* of the sum $f + F$ at each iteration. More precisely, when x_k is known, one replaces f by its linearization at this point and consider the auxiliary problem (1.11) instead of (1.10). This is also the case for other methods applied to GE (1.10), such as inexact Newton method [25] and quasi-Newton method [9]. In most of these works, the partial linearization is operated only on the single-valued part f. In their paper [39], M. Gaydu and M.H. Geoffroy proposed a local scheme for which both f and F could be approximated. This was achieved by using the concept of *set-valued differentiation* introduced by C.H.J. Pang [49]. The algorithm produces a sequence (x_k) through solving successively subproblems of the form

$$0 \in f(x_k) + Df(x_k)(x_{k+1} - x_k) + H(x_{k+1} - x_k) + F(x_k), k = 0, 1, \dots \tag{1.13}$$

where $H : X \rightrightarrows Y$ is a *positively homogeneous mapping*, which is a strictly derivative of F [39, 49] at solution $x^* \in X$ of (1.10). The key strategy of (1.13) is to take $f(x_k) + Df(x_k)(\cdot - x_k) + H(\cdot - x_k) + F(x_k)$ as an approximation for $f + F$ throughout the iterative process. Otherwise, the algorithm (1.13) requires keeping a same mapping $H(\cdot)$ for all iterations, which itself must be a set-valued derivative of F at x^*. Nevertheless, such a solution is commonly unknown in practice.

Concerning with the current work, we will propose in Chapter 4 an extension of (1.13) for solving GE (1.10). Let $\mathcal{H} : X \longrightarrow \mathcal{PH}(X, Y)$ be a given map from X to the set $\mathcal{PH}(X, Y)$ of all homogeneous mappings between X and Y. Then, at any step x_k, one updates the next iteration x_{k+1} by solving the auxiliary problem

$$0 \in f(x_k) + Df(x_k)(x - x_k) + \mathcal{H}(x_k)(x - x_k) + F(x_k). \tag{1.14}$$

It is clear that (1.13) can be subsumed as a particular case of (1.14) by letting $\mathcal{H}(x) \equiv H$. Observe that the approximating term $\mathcal{H}(x_k)(\cdot - x_k) + F(x_k)$ in (1.14) varies for each iteration. Under some assumptions on the uniformity of the "homogenous component" $\mathcal{H}(\cdot)$ as well as the metric regularity property for $Df(x_0)(\cdot) + \mathcal{H}(x_0)(\cdot)$, we prove that the scheme in (1.14) generates at least one sequence converging *R-linearly* to a solution (for *semi-local convergence*, see Theorem 4.6). The same conclusion also holds in the *local analysis*, when a solution x^* of (1.10) is considered instead of x_0. A refinement of local linear convergence is presented if a stronger estimate for differentiability at x^* is verified.

In recent decades, the idea of iterative method on *Riemannian manifolds* have been developed by many authors. These researches were motivated by problems appearing in many applications, such as, constrained optimization, singular decomposition, matrix approximations, independent component analysis, etc. For example, the eigenvalue problem can be reformulated into the form of minimizing a smooth real-valued function defined on some suitable manifolds (e.g. the *Rayleigh quotient* on *unit sphere* [1]). Regarding the survey about motivation and applications of the methods on manifolds, see the books [1, 79] and also [2, 3, 27, 43, 58, 68, 78].

Up to now, there exists a lot of results concerning the Newton-type method as well as its extension applied to smooth manifolds. For instance, the papers [8, 33, 34] extend Kantorovich's theorem, and [16] refines the Smale's theory in order to find a singularity of a smooth vector field. If f is a smooth real-valued function defined on a manifold $\mathcal{M}$, then any critical point of f is a singularity of its gradient $\operatorname{grad} f$, which solves the equation $\operatorname{grad} f(p) = 0$. Furthermore, the paper [82] considered the inclusion $0 \in f(p) + C$, where $C \subset \mathbb{R}^n$ is a cone and $f : \mathcal{M} \longrightarrow \mathbb{R}^n$ is at least C^2. Using the method introduced in [69] and the technique in [83], the authors of [82] have investigated Kantorovich's and Smale's theorems for the approximating sequence of a solution of problem $0 \in f(p) + C$.

With the motivation mentioned above, we deal with the problem of solving an inclusion of the form

$$0 \in \varphi(p) + \Phi(p), \tag{1.15}$$

in which $\varphi : \mathcal{M} \longrightarrow \mathbb{R}^n$ is a smooth map and $\Phi : \mathcal{M} \rightrightarrows \mathbb{R}^n$ has a closed graph.

Noticing that, (1.15) is reduced to the one studied in [82] by taking $\Phi(p) \equiv C$. Otherwise, by setting in (1.15) $\Phi(p) = \{0\}$ along with $\varphi = (V_1, \ldots, V_m)$, where $V_1, \ldots, V_m$ are the components of representation for the vector field V with respect to some chosen frame $\{E_1, \ldots, E_m\}$[6], the considering model (1.15) again recovers problem of singularity for vector field in the literature.
In the same spirit of (1.10), we shall make use of partial linearization in order to generate the successive approximations of a solution to (1.15). Precisely, when having an iteration p_k, let's follow a retraction R_k at the current step, and replace (1.15) by the following inclusion

$$0 \in \varphi(p_k) + \mathscr{D}\varphi(p_k)(u) + (\Phi R_k)(u), \tag{1.16}$$

to obtain a search direction u_k. Then, we update $p_{k+1} = R_k(u_k)$ as the new iteration. Thanks to a suitable choice of the set-valued part Φ and the retractions R_k, (1.16) can be viewed as a continuation of Newton's method presented in [16, 34, 82]. Following the same strategy, we prove in Chapter 5 the local and semi-local convergence of (1.16) with some assumptions imposed on the structure of the manifold as well as the behavior of the retractions R_k. Again, the notion of metric regularity property for set-valued map plays an important role in our analysis.

The whole content of this dissertation is organized as follows. The current chapter is an introduction. Chapter 2 recalls some basic backgrounds and notations which will be used throughout the thesis. Chapter 3 is devoted to the Kantorovich's and Smale's convergence theorems for generalized equations studied in [6]. Chapter 4 introduces a kind of Newton-type method solving GE (1.10) by approximating both single and multi-valued parts. In addition, we provide some results concerning the convergence of the proposed algorithm. (This chapter is based on the manuscript [7]). The next chapter takes into account a Newton-type algorithm for solving inclusions which involve set-valued maps defined on Riemannian manifolds. The main material for that chapter is based on Riemannian geometry as well as variational analysis, where metric regularity property is a key point.

[6] as knew from the basis of differential geometry, such a frame is natural

Chapter 2

Preliminaries

We recall in this chapter some preliminaries and notations that will be used throughout the thesis. The basic tools come from variational analysis and Riemannian geometry, where the metric regularity property of set-valued maps is a key concept. The author prefers to adopt the notations given in [24, 60, 61] for basic background of variational analysis and those used in [20, 75] for the notion of differential geometry.

2.1 Elementary Notations and Concepts

Unless other specifications, throughout this text, the term "space" is meant to be Banach space, which is usually denoted by upper characters X, Y, etc. The dual of X will be written as X^* while $\langle \cdot, \cdot \rangle$ will be the general duality pairing between X and X^*. For simplicity, all norm are denoted by a common notation $\|\cdot\|$, and $d(\cdot, \cdot)$ stands for the distance function. There might be no confusion when using $\|\cdot\|$, and $d(\cdot, \cdot)$ to indicate norm and distance of any Banach space. In fact, the space will be determined by the context or by the objects on which either norm or distance function acts. Functions (also, single-valued maps) are conventional written by normal character f, g while capital ones like F, G are often regarded set-valued maps. As usual, the open (closed) ball in X with center x and radius r is denoted by $\mathbb{B}_X(x, r)$ (resp. $\bar{\mathbb{B}}_X(x, r)$). When dealing with the unit balls, we write $\mathbb{B}_X$ (open) and $\bar{\mathbb{B}}_X$ (closed) respectively. In a certain situation, the space is itself clear by the context, so we frequently omit the subscripts in these notations. The basic notions of Banach space theory are assumed to be familiar, details are referred to the textbooks [28, 29] and references therein.

One may also need some set operators on Banach space. Let K and K' be two subsets of X, then their sum is defined by $K + K' = \big\{u + u' : u \in K, u' \in K'\big\}$, and $K + \emptyset = \emptyset$. If $u \in X$, then $u + K'$ represents the sum $K + K'$ with $K = \{u\}$. Furthermore, for a scalar λ and a subset $\emptyset \neq K \subset X$, the product λK is meant to be the set $\big\{\lambda u : u \in K\big\}$. K is called a cone if $\lambda K \subset K$ whenever $\lambda \geqslant 0$. If $\lambda K + (1-\lambda)K \subset K$ holds for every $\lambda \in [0,1]$, the reference set is convex.

Given now two Banach spaces X and Y, and let $f : X \longrightarrow Y$ be a (single-valued) map. If f is Fréchet differentiable at a point $x \in X$ (see [29]), then by $Df(x)$ we mean the first derivative of f at x. Otherwise, we use the notation $D^k f$ for the k-order Fréchet derivative whenever it exists. If $D^k f$ is well-define, and $v \in X$, then expression $D^k f(x)(v)^k$ stands for the value of k-linear operator $D^k f(x)$ taken at k-multiple $(v, \ldots, v) \in X^k$.

To end this section, we introduce the concept of analytic maps.

Definition 2.1 ([15])**.** A map $f : X \longrightarrow Y$ is called to be *analytic* at $x \in X$ if all derivatives $D^k f(x)$ exist, and there is a neighborhood $\mathbb{B}(x, \varepsilon)$ of x such that

$$f(y) = \sum_{k \geqslant 0} \frac{1}{k!} D^k f(x)(y-x)^k, \quad \text{for all } y \in \mathbb{B}(x,\varepsilon). \tag{2.1}$$

If f is analytic at every point of an open set U, then one says that f is analytic on U.

When f is analytic at x, then the radius of convergence for Taylor's series in the right-hand side of (2.1) can be given as follows [15]

$$R(f,x)^{-1} := \limsup_{k \to \infty} \left\| \frac{1}{k!} D^k f(x) \right\|^{1/k}. \tag{2.2}$$

2.2 Set-Valued Map, Generalized Differentiation

Throughout this dissertation, we frequently work with *set-valued maps*, which assign each element in the source space to a subset (maybe empty) of destination space. In the scope of this text, the terminologies *mapping*, *multivalued map* and *multifunction* are used as the same meaning with set-valued map while "map" is itself used for *single-valued* one. Any mapping $T : X \rightrightarrows Y$ can be identified to its graph $\operatorname{Gr} T := \big\{(x,y) \in X \times Y : y \in T(x)\big\}$. The domain $\operatorname{Dom} T$ of T is the set of all elements whose image by T is nonempty. If $\operatorname{Gr} T$ is a closed set, then we say that T is closed itself. When $\operatorname{Gr} T$ is a cone in $X \times Y$, T is called to be a positively homogeneous mapping. Sometimes, the notation $\mathcal{PH}(X,Y)$ will refer to the collection of all positively homogeneous mappings. For each $T \in \mathcal{PH}(X,Y)$, one defines its outer norm as the quantity [49, 74]

$$|T|^+ := \sup_{\|w\| \leqslant 1} \sup_{z \in T(w)} \|z\|. \tag{2.3}$$

Notice that, $|T|^+ < +\infty$ implies $T(0) = \{0\}$ (cf. [49]).

The rest of this section is left to present some concepts related to generalized differentiation for multifunctions. Firstly, we recall the notion of coderivative.

Definition 2.2 (normal cones, [60])**.** Given a nonempty subset Ω of a Banach space X.

For $\varepsilon \geqslant 0$, the ε-*normal cone* of Ω at $x \in \Omega$ is defined by

$$\widehat{N}^{\varepsilon}_{\Omega}(x) = \left\{ x^\star \in X^* : \limsup_{y \to x, y \in \Omega} \frac{\langle x^\star, y - x\rangle}{\|y - x\|} \leqslant \varepsilon \right\}, \tag{2.4}$$

and $\widehat{N}^{\varepsilon}_{\Omega}(x) := \emptyset$ for $x \notin \Omega$. In the case $\varepsilon = 0$, the corresponding cone is usually called as Fréchet normal cone $\widehat{N}_{\Omega}(x)$ to Ω at x.

The following set

$$N_\Omega(x) := \left\{ x^\star \in X^* : \exists x_k \to x, \varepsilon_k \searrow 0, x^\star_k \in \widehat{N}^{\varepsilon_k}_{\Omega}(x_k) \text{ with } x^\star_k \xrightarrow{w^*} x^\star \right\} \tag{2.5}$$

is said to be *limitting normal cone* to Ω at $x \in X$.

In (2.5), expression $x^\star_k \xrightarrow{w^*} x^\star$ means that, the sequence $x^\star_k$ converges to $x^\star$ in the weak-star topology of X^*. So, when X is finite dimension, this simply reads $x^\star_k \to x^\star$.

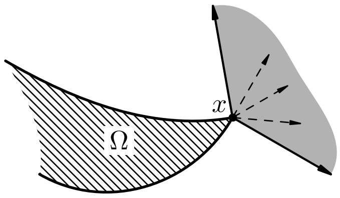
According to the Definition 2.2, it is clear that $\widehat{N}_\Omega(x) \subset N_\Omega(x)$. If the opposite inclusion is also true, which is the same as $\widehat{N}_\Omega(x) = N_\Omega(x)$, the set Ω is said to be (Clarke) *regular* at reference point x.

Definition 2.3 (coderivative)**.** Let $T : X \rightrightarrows Y$ be a given mapping between two Banach spaces X, Y and $(\bar{x}, \bar{y}) \in \operatorname{Gr} T$. The *regular (Fréchet) coderivative* $\widehat{D}^*T(\bar{x},\bar{y}) : Y^* \rightrightarrows X^*$ and *limitting coderivative* $D^*T(\bar{x},\bar{y}) : Y^* \rightrightarrows X^*$ are set-valued maps determined as follows. One has

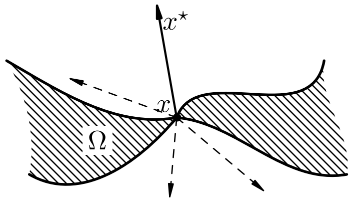
$$x^\star \in \widehat{D}^*T(\bar{x},\bar{y})(y^\star) \Longleftrightarrow (x^\star, -y^\star) \in \widehat{N}_{\operatorname{Gr} T}\big((\bar{x},\bar{y})\big) \tag{2.6}$$

and

$$x^\star \in D^*T(\bar{x},\bar{y})(y^\star) \Longleftrightarrow (x^\star, -y^\star) \in N_{\operatorname{Gr} T}\big((\bar{x},\bar{y})\big). \tag{2.7}$$



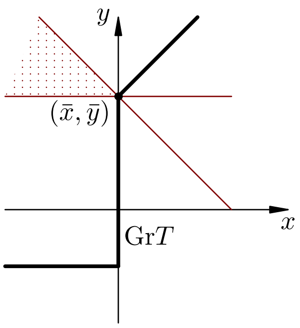
(a): Ω is Clarke regular at reference point x.



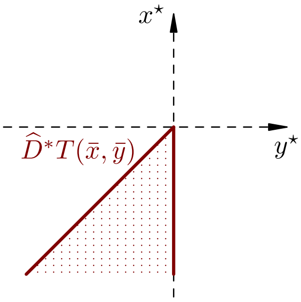
(b): Ω fails to be regular, $x^\star$ is in $N_\Omega(x)$, while $\widehat{N}_\Omega(x) = \{0\}$.

Fig. 2.1: Illustration for normal cones and Clarke regularity

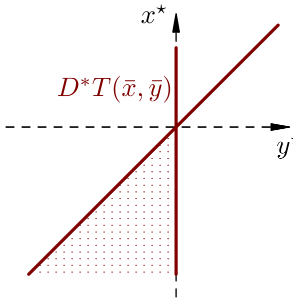
According to [60], if $f : X \longrightarrow Y$ is differentiable at $\bar{x}$ then $\widehat{D}^* f(\bar{x}) = Df(\bar{x})^*$, where A^* is the dual of linear and continuous operator A. One similar relation also holds for limitting coderivative under restricted condition that f is strictly differentiable at $\bar{x}$. Recalling strict differentiability means that $\limsup\limits_{x \neq x' \to \bar{x}} \frac{\|f(x) - f(x') - A(x - x')\|}{\|x - x'\|} = 0$ for $A = Df(\bar{x})$.



(a): Reference mapping



(b): Regular coderivative



(c): Limitting coderivative

Fig. 2.2: Example for three mappings: T, $\widehat{D}^* T(\bar{x}, \bar{y})$ and $D^* T(\bar{x}, \bar{y})$

Concerning the computation of coderivatives, let us present now a sum rule proved in [60].

Theorem 2.4. *Given a map $f : X \longrightarrow Y$ and a mapping $F : X \rightrightarrows Y$. Let $(\bar{x}, \bar{y}) \in \operatorname{Gr} F$ and suppose f is differentiable at $\bar{x}$. Then*

$$\widehat{D}^*(f + F)(\bar{x}, \bar{z})(y^\star) = Df(\bar{x})^*(y^\star) + \widehat{D}^* F(\bar{x}, \bar{y})(y^\star), \bar{z} = f(\bar{x}) + \bar{y}, \tag{2.8}$$

where $(f + F)(x) := f(x) + F(x)$. Furthermore, in the case f is strictly differentiable at $\bar{x}$, then (2.8) *is still valid with respect to the limitting coderivative.*

Finally, we end-up this section by recalling the notion of set-valued differentiability in the sense of C.H.J. Pang [49]. This will be an essential material for the developments in Chapter 4.

Definition 2.5. Consider two multivalued maps $S : X \rightrightarrows Y$ and $T : X \rightrightarrows Y$ where $T \in \mathscr{PH}(X, Y)$.

S is called to be *outer T-differentiable* at $\bar{x} \in \mathrm{Dom}(S)$ if for any $\delta > 0$ there exists a neighborhood V of $\bar{x}$ such that

$$S(x) \subset S(\bar{x}) + T(x - \bar{x}) + \delta \|x - \bar{x}\| \bar{\mathbb{B}}, \quad \text{for all} \quad x \in V. \tag{2.9}$$

S is *inner T-differentiable* at $\bar{x}$ if (2.9) is replaced by

$$S(\bar{x}) \subset S(x) - T(x - \bar{x}) + \delta \|x - \bar{x}\| \bar{\mathbb{B}}, \quad x \in V. \tag{2.10}$$

S is *T-differentiable* if it is both outer and inner T-differentiable.

We say that S is *strictly T-differentiable* at $\bar{x}$ if for any $\delta > 0$ there exists a neighborhood V of $\bar{x}$ such that

$$S(x') \subset S(x) + T(x' - x) + \delta \|x - \bar{x}\| \bar{\mathbb{B}}, \quad \text{for all} \quad x, x' \in V. \tag{2.11}$$

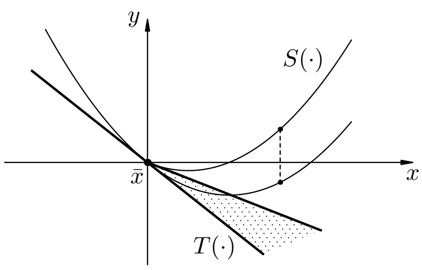


Fig. 2.3: S is T-differentiable at $\bar{x}$

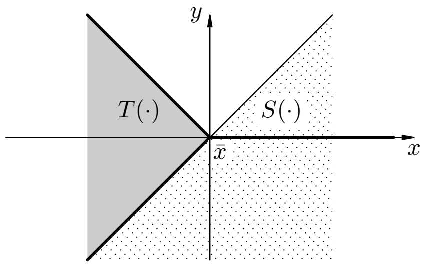


Fig. 2.4: S fails to be T-differentiable at $\bar{x}$

2.3 Lipschitz Continuity and Metric Regularity

It is well-known that a map $f : X \longrightarrow Y$ is Lipschitz continuous on a set $D \subset X$ along with a modulus κ provided that $\|f(x) - f(x')\| \leqslant \kappa \|x - x'\|$ holds for any pair x, x' in D. In order to extend such a property to a set-valued map, we need the excess as well as Hausdorff distance defined for two subsets of a space.

Definition 2.6 (Hausdorff distance)**.** Suppose K and K' are two subsets of a Banach space. The *excess* of K beyond K' is the quantity

$$e\big(K, K'\big) := \sup_{x \in K} d\big(x, K'\big) \tag{2.12}$$

with convention $e(\emptyset, K') = 0$ when $K' \neq \emptyset$ and $e(\emptyset, \emptyset) = +\infty$. Here, as usual, the distance $d(x, K')$ is given by the infimum $d(x, K') = \inf \{ \|x - u\| : u \in K' \}$.

The *Hausdorff distance* between K and K' is defined as

$$d^{\mathcal{H}}(K, K') = \max \{ e(K, K'), e(K', K) \}. \tag{2.13}$$

Alternatively, we can represent the expression (2.12) in the other form $e(K, K') = \inf \{ \tau > 0 : K \subset K' + \tau \mathbb{B} \}$. Analogously, one has (see e.g. [24]) $d^{\mathcal{H}}(K, K') = \sup_{x \in X} |d(x, K) - d(x, K')|$.

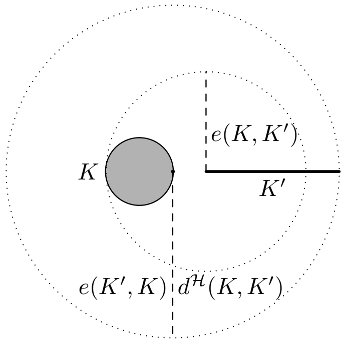


Fig. 2.5: Excess and Hausdorff distance in two dimension

Definition 2.7 (set-valued Lipschitz continuity)**.** A multifunction $S : Y \rightrightarrows X$ (here both Y and X are Banach) is said to be Lipschitz continuous on a set $\Omega \subset Y$ if there exists a constant (Lipschitz modulus) $\kappa > 0$ such that

$$d^{\mathcal{H}}(S(x), S(x')) \leqslant \kappa \|x - x'\|, \text{ for all } x, x' \in \Omega. \tag{2.14}$$

One says that $S : Y \rightrightarrows X$ has the Aubin property around $(\bar{y}, \bar{x}) \in \operatorname{Gr} S$ providing there are a constant $\kappa > 0$ along with some neighborhoods V of $\bar{y}$ and U of $\bar{x}$ such that

$$e(S(y) \cap U, S(y')) \leqslant \kappa \|y - y'\|, \text{ whenever } y, y' \in V. \tag{2.15}$$

This property is also known in the literature as pseudo-Lipschitz or Lipschitz-like. For brevity, we write $\kappa \in \operatorname{Lipmod}(S, \Omega)$ to indicate a Lipschitz modulus $\kappa > 0$ with respect to the data (S, Ω) satisfying (2.14). In addition, we also denote $\operatorname{Lip} S(\bar{y}, \bar{x})$ the infimum of all parameter $\kappa > 0$ for which (2.15) holds under some neighborhoods V of $\bar{y}$ and U of $\bar{x}$. Conventionally, we set $\operatorname{Lip} S(\bar{y}, \bar{x}) = +\infty$ in the case (2.15) is absent.

It is evident to see that, the notion of Lipschitz continuity defined by (2.14) covers the usual one for a single-valued case. On the other hand, pseudo-Lipschitz property was also known to be equivalent to another important concept which is presented below.

Definition 2.8 (metric regularity, [44])**.** Let $T : X \rightrightarrows Y$ be a given mapping. T is called to be *metrically regular* on a set $\mathcal{V} \subset X \times Y$ with a modulus $\tau > 0$ if

$$(x, y) \in \mathcal{V} \Longrightarrow d\big(x, T^{-1}(y)\big) \leqslant \tau d\big(y, T(x)\big). \tag{2.16}$$

The mapping T is (locally) metrically regular around $(\bar{x}, \bar{y}) \in \operatorname{Gr} T$ if there exists $\tau > 0$ so that (2.16) holds in the case $\mathcal{V}$ is a neighborhood of $(\bar{x}, \bar{y})$. Infimum of all such moduli $\tau > 0$ is denoted by $\operatorname{Reg} T(\bar{x}, \bar{y})$.

For shortness, let's write $\tau \in \operatorname{Regmod}(T, \mathcal{V})$ to indicate the property (2.16). Also, the infimum of all $\tau > 0$ for which (2.16) fulfills will be denoted by $\mathbf{Reg}_{\mathcal{V}}(T)$.

Local metric regularity property of a mapping T around a point $(\bar{x}, \bar{y})$ implies the validity of pseudo-Lipschitz continuity for $S = T^{-1}$ at $(\bar{y}, \bar{x})$ and vice versa. More precisely, one has (see [24, 44, 60])

$$\operatorname{Reg} T(\bar{x}, \bar{y}) = \operatorname{Lip} T^{-1}(\bar{y}, \bar{x}). \tag{2.17}$$

Ultimately, we discuss a well-known characterization of regularity modulus through coderivative. According to the complication of its full proof, it will be skipped here and left to the references e.g. [59, 74].

Theorem 2.9 (Mordukhovich criterion)**.** *For a closed multifunction $\Phi : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and a pair $(\bar{x}, \bar{y})$ in $\operatorname{Gr} \Phi$, then Φ is metrically regular around $(\bar{x}, \bar{y})$ if and only if*

$$0 \in D^*\Phi(\bar{x}, \bar{y})(y^\star) \Longrightarrow y^\star = 0, \tag{2.18}$$

or equivalently, $|D^*\Phi^{-1}(\bar{y}, \bar{x})|^+ < +\infty$. In such a case, one has $\operatorname{Reg} \Phi(\bar{x}, \bar{y}) = |D^*\Phi^{-1}(\bar{y}, \bar{x})|^+$.

In terms of normal cone, Mordukhovich criterion can be reformulated as (see an illustration shown in Figures 2.6 and 2.7)

$$\operatorname{Reg} \Phi(\bar{x}, \bar{y}) < +\infty \Longleftrightarrow (\{0\} \times \mathbb{R}^n) \cap N_{\operatorname{Gr} \Phi}(\bar{x}, \bar{y}) = \{(0, 0)\}. \tag{2.19}$$

A similar criterion treating with *semi-local*[1] modulus of regularity around any point has been investigated in the work [59]. The readers are referred to [62] for

[1]the terminology semi-local is due to [60]

complete characterization of metric regularity property in infinitely dimensional case. As a direct consequence, a C^1 map $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ admits the metric regularity property at $\bar{x}$ if its derivative $Df(\bar{x})$ is surjective, i.e., the Jacobian $\mathrm{Jac}_f(\bar{x})$ has full-row rank (Graves theorem [24, Chapter 5]).

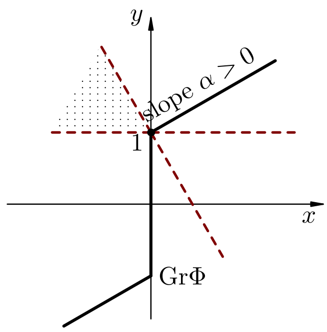


Fig. 2.6: Validity of metric regularity

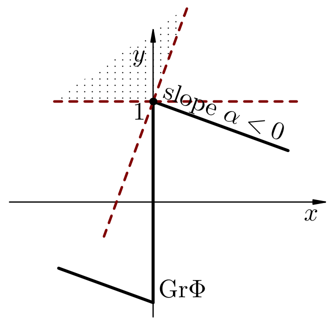


Fig. 2.7: Failure of metric regularity

2.4 Backgrounds from Riemannian Geometry

This section is devoted to some basic backgrounds of differential geometry which will be needed for the developments of Chapter 5. With the goal of unifying the notations, the manifolds are written by calligraphic uppercases like $\mathcal{M}$, $\mathcal{N}$, etc. The term "map" always indicates single-valued map defining on some manifold, and "function" is meant to be a map takes value in the real line $\mathbb{R}$. In addition, the word "smooth" will be differentiable up to a necessary order (at least C^1).

2.4.1 Fundamentals on Smooth Manifolds

Definition 2.10 (smooth manifold,[20, 75])**.** Let $\mathcal{M}$ be a nonempty set. It is a smooth manifold (or variety) of dimension m if there exists a family of injective maps $\mathbf{x}_\alpha : U_\alpha \longrightarrow \mathcal{M}$ where U_α is an open subset of $\mathbb{R}^m$ such that:

(i) $\bigcup_\alpha U_\alpha = \mathcal{M}$;

(ii) for any α and β, if $W = \mathbf{x}_\alpha(U_\alpha) \cap \mathbf{x}_\beta(U_\beta) \neq \emptyset$, then both $\mathbf{x}_\alpha^{-1}(W)$ and $\mathbf{x}_\beta^{-1}(W)$ are open sets in $\mathbb{R}^m$, and the composition $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ is smooth;

(iii) the family $\big(\mathbf{x}_\alpha, U_\alpha\big)$ is maximal relative to properties (i) and (ii).

Such a collection $\big\{\big(\mathbf{x}_\alpha, U_\alpha\big)\big\}$ satisfies (i) and (ii) is called a smooth structure on $\mathcal{M}$. If $p \in \mathbf{x}_\alpha(U_\alpha)$, then the pair $\big(\mathbf{x}_\alpha, U_\alpha\big)$ (or $\mathbf{x}_\alpha$ for shortness) is a parametrization (equivalently, a system of coordinate) of $\mathcal{M}$ at p. Smooth structure forms a topology on $\mathcal{M}$ (see [20]). A few important examples for smooth manifolds are vector spaces; m-unit sphere $\mathbb{S}^m$, orthogonal group O_m [1]; manifold of positively definite matrices Pos_m [11, 55]; the real $\mathbb{R}P^m$ projective spaces [20]; the Lie groups [57].

Definition 2.11 (smooth maps,[20])**.** Given two smooth manifolds $\mathcal{M}$ and $\mathcal{N}$. A map $\varphi : \mathcal{M} \longrightarrow \mathcal{N}$ is said to be smooth at $p \in \mathcal{M}$ provided that for each parametrization $\mathbf{y} : V \longrightarrow \mathcal{N}$ of $\mathcal{N}$ at $\varphi(p)$, there exists a corresponding one $\mathbf{x} : U \longrightarrow \mathcal{M}$ of $\mathcal{M}$ at p such that the composition $\mathbf{y}^{-1} \circ \varphi \circ \mathbf{x}$ is well-defined on a neighborhood of $\mathbf{x}^{-1}(p)$ and is smooth at that point.

Hence, a function $f : \mathcal{M} \longrightarrow \mathbb{R}$ is smooth at p if for some local coordinate $\mathbf{x} : U \longrightarrow \mathcal{M}$ with $p \in \mathbf{x}(U)$ one has $f \circ \mathbf{x}$ is smooth (by considering the natural smooth structure $(\mathrm{id}, \mathbb{R})$ on $\mathbb{R}$). We will denote by $\mathcal{F}(p)$ the collection of all functions defined on a neighborhood of p and smooth at p. Similarly, notation $\mathcal{F}(\Omega)$ indicates the set of all smooth functions whose domains contain Ω.

Definition 2.12 (smooth curve and tangent vector)**.** A *curve* on some variety $\mathcal{M}$ is one *continuous* map $\gamma : I \longrightarrow \mathcal{M}$, where I is an open interval of the real line $\mathbb{R}$. If $[a, b] \subset I$, then the restriction of γ on $[a, b]$ is a *path* (or sometimes *segment*) joining $\gamma(a)$ to $\gamma(b)$. γ is called a smooth curve if it is smooth at every point of I.

A *tangent vector* (or *velocity*) of smooth curve $\gamma : I \longrightarrow \mathcal{M}$ at $t_0 \in I$ is a derivation $\gamma'(t_0)$ defined by the following rule

$$\gamma'(t_0)f := \frac{d}{dt}(f \circ \gamma)\Big|_{t=t_0}, \ \forall f \in \mathcal{F}(\mathcal{M}). \tag{2.20}$$

A tangent vector to $\mathcal{M}$ at p is defined as $\gamma'(0)$ for some smooth curve γ on $\mathcal{M}$ with $\gamma(0) = p$. The set of all such vectors is the *tangent space* $T_p\mathcal{M}$ of $\mathcal{M}$ at p (see Fig. 2.8).

Sometimes, the velocity γ' is also denoted as $\dot{\gamma}$ or $\frac{d\gamma}{dt}$. All tangent spaces are vectorial of the same dimension as the original manifold $\mathcal{M}$.

Definition 2.13 (tangent bundle and vector field)**.** For a given manifold $\mathcal{M}$ of dimension m, its *tangent bundle* $T\mathcal{M}$ can be viewed as the disjoint union of all tangent spaces $T\mathcal{M} := \big\{(p, v_p) : v_p \in T_p\mathcal{M}\big\}$. $T\mathcal{M}$ endowed a natural smooth structure with dimension $2m$ (cf. [20, 56]).

A *vector field* on $\mathcal{M}$ is a smooth map $V : \mathcal{M} \longrightarrow T\mathcal{M}$ so that $V(p) \in \{p\} \times T_p\mathcal{M}$, i.e., its valued at p is a tangent vector at the reference point p.

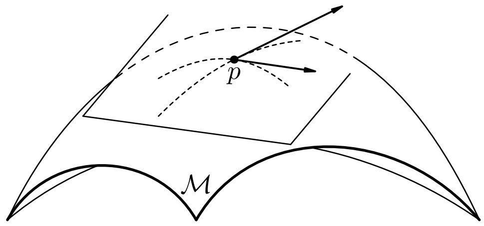


Fig. 2.8: An illustration for the tangent space

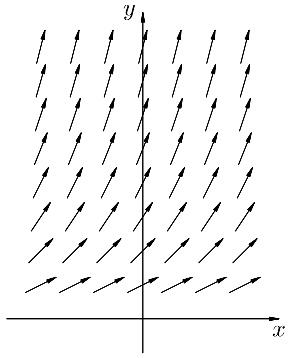


Fig. 2.9: Example of a vector field

To avoid some undesired complexity, we usually write $V_p \in T_p\mathcal{M}$ as the value of V at p so that $V(p) = (p, V_p)$. We also denote by $\mathcal{V}(\Omega)$ the set of all smooth maps defined on $\Omega \subset \mathcal{M}$ whose values are in the tangent bundle of $\mathcal{M}$.

Given now $V \in \mathcal{V}(\Omega)$ and $f \in \mathcal{F}(\Omega)$. The action of V on f is a function Vf defined on Ω as follows. For each $p \in \Omega$, the value of Vf at p is V_pf. Recall that $V_pf \in \mathbb{R}$ makes sense due to Definition 2.12, since $V_p \in T_p\mathcal{M}$.

Let W be another smooth vector field and $f \in \mathcal{F}(\Omega)$, then we can apply V to the function Wf and obtain a new object VW by the rule $(VW)f := V(Wf)$. Finally, the *Lie bracket* $[V, W]$ of V and W is a vector field defined by $[V, W] = VW - WV$.

We finish this subsection by recalling the differential of a smooth map.

Definition 2.14. Let $\varphi : \mathcal{M} \longrightarrow \mathcal{N}$ be a map which is smooth at p. The differential of φ at p, written as $d\varphi_p$ (or also $d\varphi(p)$), is a linear map from $T_p\mathcal{M}$ into $T_{\varphi(p)}\mathcal{N}$. This map is given by the formula $[d\varphi_p(v)]f := v(f \circ \varphi)$, $v \in T_p\mathcal{M}$, $f \in \mathcal{F}(\mathcal{N})$. Alternatively, taking a smooth curve γ with $\gamma(0) = p$ and $\gamma'(0) = v$, then $d\varphi_p(v) = (\varphi \circ \gamma)'(0)$.

2.4.2 Riemannian Metric, Covariant Derivative and Parallelism

Until now, we have dealt with basic objects on Riemannian geometry. Firstly, we discuss the metric structure on a manifold.

Definition 2.15 (Riemannian metric)**.** A *Riemannian metric* g on $\mathcal{M}$ is a correspondence which assigns to each $p \in \mathcal{M}$ an inner product $\langle \cdot, \cdot \rangle_p$ on $T_p\mathcal{M}$ varies smoothly in p by the sense described as follows. Let $\mathbf{x} : U \longrightarrow \mathcal{M}$ be a local coordinate at p, and $\partial_i = \frac{\partial}{\partial x_i}$ be the vector field such that $(\partial_i)_q f = \frac{\partial (f \circ \mathbf{x})}{\partial x_i}(\mathbf{x}^{-1}(q))$. Then all maps $g_{i,j}(u) := \left\langle (\partial_i)_{\mathbf{x}(u)}, (\partial_j)_{\mathbf{x}(u)} \right\rangle_{\mathbf{x}(u)}$ are smooth on U.

A manifold endowed with a Riemannian metric becomes a Riemannian manifold (of the same dimension). Here and in the sequel, the term "manifold" always refers to Riemannian sense. For a given smooth path $\gamma : [a, b] \longrightarrow \mathcal{M}$, one defines its arc length (shortly, length) as the quantity

$$\ell(\gamma) = \int_a^b \|\gamma'(t)\|_{\gamma(t)}\, dt = \int_a^b \langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)}^{1/2}\, dt, \tag{2.21}$$

where $\|\cdot\|_z = \langle \cdot, \cdot \rangle_z^{1/2}$ stands for the norm induced on $T_z\mathcal{M}$. If such a path γ is just piecewise smooth, i.e. there is an partition $t_0 = a < t_1 < \ldots < t_k = b$ in I for which $\gamma|_{(t_i, t_{i+1})}$ is smooth, then $\ell(\gamma)$ is the sum taken over all components $\ell\left(\gamma|_{[t_i, t_{i+1}]}\right)$. The distance function on $\mathcal{M}$ will be defined as follows

$$d_{\mathcal{R}}(p, q) := \inf \Big\{ \ell(\gamma) : \gamma \text{ is a piecewise smooth path connecting } p \text{ to } q \Big\}. \tag{2.22}$$

It is well-known in the literature that $\mathcal{M}$ endowed with distance $d_{\mathcal{R}}(\cdot, \cdot)$ is a metric space whose topology coincides with the initial topology of the variety [55, 56].

Definition 2.16 (connection)**.** An *affine connection* ∇ is a map $\nabla : \mathcal{V}(\mathcal{M}) \times \mathcal{V}(\mathcal{M}) \longrightarrow \mathcal{V}(\mathcal{M})$ which sends a pair of vector fields (X, Y) into another one $\nabla(X, Y) := \nabla_X Y$ satisfying the three conditions below:

(i) $\nabla_{f_1 X_1 + f_2 X_2} Y = f_1 \nabla_{X_1} Y + f_2 \nabla_{X_2} Y$; $X_j, Y \in \mathcal{V}(\mathcal{M})$, $f_j \in \mathcal{F}(\mathcal{M})$

(ii) $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$; $X, Y, Z \in \mathcal{V}(\mathcal{M})$

(iii) $\nabla_X (fY) = (Xf)Y + f \nabla_X Y$; $X, Y \in \mathcal{V}(\mathcal{M})$, $f \in \mathcal{F}(\mathcal{M})$.

Here the operator fY means a vector field that $(fY)_p = f(p) Y_p$. Additionally, ∇ is said to be a Riemannian (or, Levi-Civita) connection if it is symmetric (i.e. $[X, Y] \equiv \nabla_X Y - \nabla_Y X$) and is compatible with the Riemannian metric

$$X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle. \tag{2.23}$$

In what follows, all connection under consideration are always Riemannian connection. This kind of concept plays an essential roles for further analysis later. In particular, it allows us to develop the covariant theory for vector fields and for functions as an extension of usual differentiation on Euclidean spaces. A vector field V along a curve $\gamma : I \longrightarrow \mathcal{M}$ is a correspondence assigning each $t \in I$ to $V(t) \in T_{\gamma(t)}\mathcal{M}$. One says that a vector field V is smooth iff any map $f_V(t) := V(t) f$

is smooth for all $f \in \mathcal{F}(\mathcal{M})$. The set of all such objects is denoted by $\mathcal{V}(\gamma)$.

Proposition 2.17 (covariant derivative of vector field,[20, 56])**.** *Given a smooth curve $\gamma : I \longrightarrow \mathcal{M}$. There exists a (unique) operator $\frac{D^\gamma}{dt}$ assigning $V \in \mathcal{V}(\gamma)$ to $\frac{D^\gamma}{dt}(V) = \frac{D^\gamma V}{dt} \in \mathcal{V}(\gamma)$ such that:*

(i) $\frac{D^\gamma}{dt}(aV + bW) = a\frac{D^\gamma}{dt}(V) + b\frac{D^\gamma}{dt}(W)$; $V, W \in \mathcal{V}(\gamma)$, $a, b \in \mathbb{R}$.

(ii) $\frac{D^\gamma}{dt}(fV) = f'V + f\frac{D^\gamma}{dt}(V)$; $V \in \mathcal{V}(\gamma)$, $f : I \longrightarrow \mathbb{R}$ is smooth.

(iii) If $V(t) = Y_{\gamma(t)}$ for $Y \in \mathcal{V}(\mathcal{M})$ then $\frac{D^\gamma}{dt} = \nabla_{\gamma'}Y$.

The field $\frac{D^\gamma V}{dt}$ in Proposition 2.17 is the *covariant derivative* of V along γ. It allows us to define the notion of parallellism along a curve as in the next definition.

Definition 2.18 (parallelism)**.** A vector field V along the curve $\gamma : I \longrightarrow \mathcal{M}$ is *parallel* if its covariant derivative is vanishing $\frac{D^\gamma}{dt}V \equiv 0$. For $a, b \in I$, the associated *parallel transport* $P_\gamma^{a,b} : T_{\gamma(a)}\mathcal{M} \longrightarrow T_{\gamma(b)}\mathcal{M}$ along γ is a map determined as follows. If $v \in T_{\gamma(a)}\mathcal{M}$, then the initial value problem

$$V(a) = v, \frac{D^\gamma V}{dt}(t) = 0, V \in \mathcal{V}(\gamma) \tag{2.24}$$

has a unique solution (see [56]), and one sets $P_\gamma^{a,b}(v) := V(b)$, where V is solution of the system (2.24).

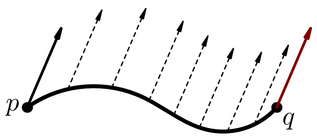


Fig. 2.10: A parallel field and the corresponding transportation

A trivial example for parallel transformation is the usual translation $T_x(v) := x + v$ in any Euclidean space. For a given curve γ, $P_\gamma^{a,b}$ is always a linear isometry from $T_{\gamma(a)}\mathcal{M}$ into $T_{\gamma(b)}\mathcal{M}$ whenever it exists. Otherwise, the relation $P_\gamma^{a,b} = \left(P_\gamma^{b,a}\right)^{-1}$ holds if either $P_\gamma^{a,b}$ or $P_\gamma^{b,a}$ makes sense. In certain situations, when either γ or parameters a, b are specified from the context, we ignore the appearance of such objects in the notation of parallel transport.

Finally, we describe a short survey of covariant derivation corresponding to a smooth map which takes value in an Euclidean space. Let f be a smooth function, one defines its covariant derivative $\mathscr{D}f$ by setting

$$\mathscr{D}f(p)(u) = (\nabla_X f)(p) := (Xf)(p), \quad u \in T_p\mathcal{M}, \tag{2.25}$$

for $X \in \mathcal{V}(\mathcal{M})$ such that $X_p = u$. The gradient $\operatorname{grad} f(p)$ of f at p is a vector given by

$$\langle \operatorname{grad} f(p), u\rangle := (Xf)(p),\ X \in \mathcal{V}(\mathcal{M}),\ X_p = u. \tag{2.26}$$

Let $F = (f_1, \ldots, f_m) : \mathcal{M} \longrightarrow \mathbb{R}^n$ be a smooth map. The covariant derivative $\mathscr{D}F(p)$ of F at $p \in \mathcal{M}$ satisfies the expression $\mathscr{D}F(p)(u) := \Big(\mathscr{D}f_1(p)(u), \ \cdots \ , \mathscr{D}f_n(p)(u)\Big)$. The map $\mathscr{D}F(p)$ is a linear transformation from $T_p\mathcal{M}$ into $\mathbb{R}^m$ [16].

2.4.3 Geodesic, Retraction, Vector Transportation

Definition 2.19 (geodesic)**.** Let $\gamma : I \longrightarrow \mathcal{M}$ be a given smooth curve on a manifold $\mathcal{M}$. γ is said to be a geodesic if and only if its velocity is parallel along γ itself, i.e., $\nabla_{\gamma'}\gamma' \equiv 0$.

On $\mathbb{R}^m$ endowed with Euclidean metric, any geodesic is just a straight line. On the unit sphere $\mathbb{S}^2$ with the metric inherited from Euclidean distance, then geodesic is one of its *great circles*.

It is well-known that, for each pair (p, u) of the tangent bundle, there exists a geodesic passes through p with velocity u which is defined on an open interval of the real line numbers. If such a geodesic can be extended onto the whole $\mathbb{R}$, we say that it is a geodesic line. The manifold having property that all geodesics are defined on $\mathbb{R}$ is said to be geodesically complete (shortly, complete). Due to the Hopf-Rinow theorem [20, 56, 75], such a situation takes place if the corresponding metric space $\big(\mathcal{M}, d_{\mathscr{R}}\big)$ is complete (and vice versa). In the scope of this dissertation, all manifolds under consideration are assumed to be complete.

Definition 2.20 (exponential map)**.** Given a complete Riemannian manifold $\mathcal{M}$ of finite dimension. The *exponential map* is defined by sending $(p, v) \in T\mathcal{M}$ into $\operatorname{Exp}(p, u) = \gamma(p, u, 1) \in \mathcal{M}$, where $\gamma(p, u, \cdot)$ stands for the geodesic on $\mathcal{M}$ going through p at the instance $t = 0$ with velocity $\frac{d\gamma(p,u,\cdot)}{dt}(0) = u$.

For $q \in \mathcal{M}$, the exponential $\exp_q(\cdot)$ at q is the restriction of Exp onto tangent space $T_q\mathcal{M}$. More precisely, $\exp_q(v) := \operatorname{Exp}(q, v)$ for $v \in T_q\mathcal{M}$.

The exponential maps are very important objects (for instance, smooth curves minimizing the arc length are geodesic [20]). At any point p of $\mathcal{M}$ there exists a normal neighborhood which is convex [20, 55]. Convexity means that, for any pair of points (p, q) there is a geodesic joining p to q. Alternatively, a subset $\Omega \subset \mathcal{M}$ is said to be strongly convex if the minimizing geodesic linking two points p and q is contained in Ω whenever $p \in \Omega$ and $q \in \Omega$.

Although they have many fine properties, the exponential maps might be often expensive to compute in practice due to the complexity of solving ordinary differential equations on manifolds. Instead of that, one can consider some other replacing objects named as retraction. In fact, most of the developments presented in Chapter 5 are based on those.

Definition 2.21 (retraction,[1])**.** Let p be in a manifold $\mathcal{M}$. Retraction at p is a smooth map R_p from the tangent space $T_p\mathcal{M}$ into $\mathcal{M}$ itself such that:
(i) $R_p(0_p) = p$, where 0_p is the origin of $T_p\mathcal{M}$;
(ii) under the canonical identification $T_{0_p}\big(T_p\mathcal{M}\big) \simeq T_p\mathcal{M}$, one has

$$\big(dR_p\big)(0_p) = \mathrm{id}_{T_p\mathcal{M}}\,. \tag{2.27}$$

If R_p is defined for every p in $\mathcal{M}$, we call $R : (p, u) \longmapsto R_p(u)$ is the retraction of tangent bundle.

Condition (ii) in Definition 2.21 can be seen as local rigidity condition (see more in [1]). Together with (i), it tells us that, for each $u \in T_p\mathcal{M}$, the correspondence $t \longmapsto R_p(tu)$ forms a smooth curve passing to p with velocity u. Rigidity permits us to establish the local property below, which is motivated from the existence of normal neighborhoods above.

Proposition 2.22 (retraction normal pair)**.** *Suppose that the domain of retraction R contains a set of the form $\big\{(q, v) : q \in \Omega, v \in T_q\mathcal{M}\big\}$, where $\Omega \subset \mathcal{M}$ is open. Then, there exists a pair of real-valued functions $\lambda_R, \iota_R : \Omega \longrightarrow (0, +\infty)$ for which the following statements hold. Given $p \in \Omega$, and $d_{\mathscr{R}}(q, p) < \lambda_R(p)$, then the map R_q is injective in the ball $\iota_R(p)\mathbb{B}_q$ of the tangent space $T_q\mathcal{M}$. In addition, the ball in $\mathcal{M}$ with center p and radius $\lambda_R(p)$ is contained into the image $R_q\left(\iota_R(p)\mathbb{B}_q\right)$ for every q with $d_{\mathscr{R}}(p, q) < \lambda_R(p)$.*

Here, $\mathbb{B}_q$ stands for the open unit ball of the tangent space $T_q\mathcal{M}$ associated with the norm $\|\cdot\|_q$ induced by Riemannian metric on $\mathcal{M}$. A pair of functions in Proposition 2.22 is said to be a *R-normal pair* (shortly, normal pair) for Ω.

Proof. Fix $p \in \Omega$ and let $\mathbf{x} : U \longrightarrow \mathcal{M}$ be a local coordinate at $p = \mathbf{x}(\bar{z})$. Then, it is possible to use $(z, z) \in U \times U \longmapsto (\mathbf{x}(z), \mathbf{x}(z))$ as a local coordinate of $\mathcal{M} \times \mathcal{M}$ at (p, p). Consider the map given by $F(q, v) := (q, R_q(v))$, $q \in \Omega$, $v \in T_q\mathcal{M}$. We have $F(p, 0_p) = (p, p)$, and by virtue of the rigidity condition, the Jacobian matrix of F with respect to the parametrization above at $(p, 0_p) \in T\mathcal{M}$ can be represented as follows

$$\begin{pmatrix} I & 0 \\ * & I \end{pmatrix}.$$

Hence, the inverse mapping theorem is applicable. Based on this fact, we can follow the arguments of proving the existence theorem for normal neighborhood into our situation here. Because of this analogousness, the details should be omitted, and we refer to [56, Lemma 5.12] and [20, Theorem 3.7] for full arguments. □

It is easy to see that the exponential map is of course a retraction. Another general class of retractions which is generated from variation of exponential map was introduced in [68]. Those have a very interesting property, which represents the parallel transport in terms of differential of retraction. According to [1] and

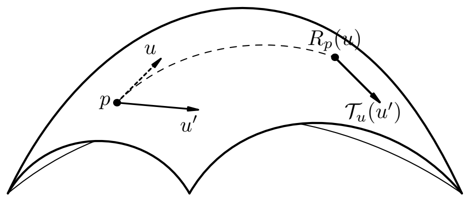


Fig. 2.11: Vector transportation on manifold

[68], it might be reasonable to introduce the wide collection of transformations, that is, the so-called vector transport. An abstract transport is a map $\mathcal{T}_{(\cdot)}(\cdot)$ defined on the Whitney sum $T\mathcal{M} \oplus T\mathcal{M}$ satisfying some certainly complement conditions. Up to the current text, we at most handle only the ones induced by the differential of a retraction. The next definition is in this sense.

Definition 2.23 (differentiated transportation)**.** Given a retraction R_p and let q be in $\mathcal{M}$. We defines a vector transport $\mathcal{T}_R^{p,q} : T_p\mathcal{M} \longrightarrow T_q\mathcal{M}$ by

$$\mathcal{T}_R^{p,q}(w) = (dR_p)_{\bar{u}}(w) = \frac{d}{dt}\Big\{R_p(\bar{u} + tw)\Big\}_{t=0}, \quad \text{for } R_p(\bar{u}) = q. \tag{2.28}$$

If it has at least two vectors for $\bar{u}$ in (2.28), then the chosen element will be specified clearly in the context.

Chapter 3

Josephy-Newton Method under Kantorovich's and Smale's Approaches

The current chapter deals with the convergence of *Josephy-Newton method* (JNm) applied to generalized equations with assumptions of type Kantorovich and Smale. Kantorovich-type convergence analyses for solving inclusions was studied by many authors e.g. in [21, 67], whereas it seems to have very few papers which extend the Smale's theories to this universal context. The main results of this chapter concentrate on four convergent theorems in Section 3.2 and Section 3.3. Recall that a generalized equation (GE) is of the form

$$0 \in f(x) + F(x), \tag{3.1}$$

while the JNm with respect to (3.1) is represented by recurrent procedure

$$0 \in f(x_k) + Df(x_k)(x_{k+1} - x_k) + F(x_{k+1}), k = 0, 1, \ldots \tag{3.2}$$

A typical property of the scheme described in (3.2) is that, *not all* resulting sequences are convergent. For instance, Figure 3.1 shows a simple situation where (3.2) may produce a divergent *Josephy-Newton sequence* (JNseq). Nevertheless, if one of those converges, then the corresponding limit will be a solution of problem (3.1). This is due to the closedness of $\mathrm{Gr}(F)$ which preserves the inclusion after letting the limit in (3.2).

At the beginning, let us examine a few results concerning the stability of metric regularity which play an essential role in our results.

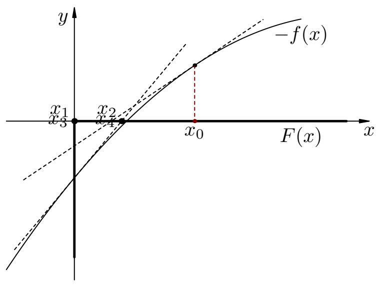


Fig. 3.1: A Josephy-Newton sequence does not converge

3.1 Stability of Metric Regularity

The first theorem of this section is stated as follows.

Theorem 3.1 (local stability)**.** *Given two Banach spaces X and Y, and let $\Phi : X \rightrightarrows Y$ be a closed mapping. Suppose $(\bar{x}, \bar{y}) \in \mathrm{Gr}(\Phi)$ and $\kappa \in \mathrm{Regmod}\,(\Phi, V_{a,b})$, where $V_{a,b} = \mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}, b)$. Let $\delta > 0$, $L \in (0, \kappa^{-1})$, and set $\tau = \kappa/(1 - \kappa L)$. Considering some positive constants α, β such that*

$$2\alpha + \beta\tau < \min\{a, \delta/2\}, \beta(\tau + \kappa) < \delta, 2c\alpha + \beta(1 + c\tau) < b, \tag{3.3}$$

with $c := \max\{1, \kappa^{-1}\}$. If $g : X \longrightarrow Y$ is a map satisfying $L \in \mathrm{Lipmod}(\Omega, g)$ for $\Omega = \mathbb{B}(\bar{x}, \delta)$, and the sum $\Psi = \Phi + g$ has closed graph, then one has $\tau \in \mathrm{Regmod}\,(\Psi, V_{\alpha,\beta})$, where $V_{\alpha,\beta} = \mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(\bar{y} + g(\bar{x}), \beta)$.

The next theorem concerns with semi-local metric regularity property. For the convenience of reading, we introduce a useful notation. Corresponding to a given mapping $T : X \rightrightarrows Y$, a point $\bar{x} \in X$ and two constants $r > 0$, $s > 0$ we define

$$V(T, \bar{x}, r, s) := \big\{(x, y) \in X \times Y : \|x - \bar{x}\| \leqslant r, d\big(y, T(x)\big) < s\big\}. \tag{3.4}$$

Using this notation, we are now in position to assert the result mentioned above.

Theorem 3.2 (semi-local stability)**.** *Let $\Phi : X \rightrightarrows Y$ be a given closed mapping, and let $\bar{x} \in \mathrm{Dom}(\Phi)$. Let $r > 0$, $s > 0$, and suppose $\kappa > 0$ being such that $\kappa \in \mathrm{Regmod}\,\big(\Phi, V(\Phi, \bar{x}, r, s)\big)$. For some $L \in (0, \kappa^{-1})$, we set $\tau = \kappa/(1 - \kappa L)$. If $g : X \longrightarrow Y$ is Lipschitz continuous on $\bar{\mathbb{B}}(x, r)$ with a modulus L, and $\Psi := \Phi + g$ has closed graph, then it holds that $\tau \in \mathrm{Regmod}\,\big(\Psi, V(\Psi, x_0, r', s')\big)$, with $r' = r/4$ and $s' = \min\big\{s, \frac{r}{5\tau}\big\}$.*

Both Theorems 3.1 and 3.2 can be subsumed as particular cases of the ones proved in [5]. Precisely, we just apply [5, Theorem 3.2] and [5, Theorem 6.2] with respect to a single-valued perturbation g. In the rest of this section, the author would suggest some developments based on two preceding theorems which follow a little different approach from [5, Theorem 3.2] and [5, Theorem 6.2].

Theorem 3.3 (local stability revisited)**.** *Given two closed mappings $\Phi : X \rightrightarrows Y$ and $G : X \rightrightarrows Y$ from X to Y (both of them are Banach spaces). Assume that $(\bar{x}, \bar{y}) \in \mathrm{Gr}(\Phi)$ and $\mathcal{V} = \mathbb{B}(\bar{x}, r) \times \mathbb{B}(\bar{y}, s)$ is a neighborhood on which Φ is metrically regular with a modulus $\kappa > 0$. Consider some constants $\eta > 0$, $L > 0$, $r' > 0$ and $s' > 0$ satisfying*

$$\begin{cases} \kappa L < 1, \\ \frac{2}{1-\kappa L} r' + \frac{\kappa}{1-\kappa L} s' + \frac{\kappa}{1-\kappa L} \eta < r, \\ \frac{2L}{1-\kappa L} r' + \frac{1}{1-\kappa L} s' + \frac{1}{1-\kappa L} \eta < s. \end{cases} \tag{3.5}$$

Let $\bar{z} \in G(\bar{x})$ be such that $e\big(G(\bar{x}), \bar{z}\big) \leqslant \eta$. If $L \in \mathrm{Lipmod}\big(G, \mathbb{B}(\bar{x}, r)\big)$ and the sum $\Psi = \Phi + G$ has closed graph, then Ψ is metrically regular on neighborhood $\mathcal{V}' = \mathbb{B}(\bar{x}, r') \times \mathbb{B}(\bar{y} + \bar{z}, s')$ together with a modulus $\tau = \frac{1}{1-L\kappa}\kappa$.

Suppose $a = r$, $b = s$, α, β and δ fulfill (3.3), then for $\eta > 0$ small enough, the constraint (3.5) holds with $r' = (1 - \kappa L)\alpha$, $s' = \beta$. Contrarily, when $r = \min\big\{a, \delta/2\big\}$, $s = b$, r' and s' obey (3.5), then $\alpha = \frac{1}{2}\min\{r', c^{-1}Lr'\}$ and $\beta = \frac{1}{2(1+c\tau)} s'$ satisfy (3.3). Thus, the two Theorems 3.1 and 3.3 can be subsumed together.

Proof. Let $\kappa' > \kappa$ and $L' > L$ such that

$$\begin{cases} \kappa' L' < 1, \\ \frac{2}{1-\kappa' L'} r' + \frac{\kappa}{1-\kappa' L'} s' + \frac{\kappa}{1-\kappa' L'} \eta < r, \\ \frac{2L'}{1-\kappa' L'} r' + \frac{1}{1-\kappa' L'} s' + \frac{1}{1-\kappa' L'} \eta' < s. \end{cases} \tag{3.6}$$

Pick $(x, y) \in \mathcal{V}'$ and suppose $d\big(y, \Psi(x)\big) > 0$ (the other case is trivial). Let $R > d\big(y, \Psi(x)\big)$ and take a point $w_0 \in G(x)$ with $d\big(y - w_0, \Phi(x)\big) < R$. By setting $x_0 = x$ and $y_0 = y - w_0$, we claim $(x_0, y_0) \in \mathcal{V}$. In fact, it is sufficient to verify $\|y_0 - \bar{y}\| < s$ only. We have

$$\begin{aligned} \|y_0 - \bar{y}\| &= \|y - (\bar{y} + \bar{z}) - (w_0 - \bar{z})\| \leqslant \|y - (\bar{y} + \bar{z})\| + \|w_0 - \bar{z}\| \\ &\leqslant \|y - (\bar{y} + \bar{z})\| + d\big(w_0, G(\bar{x})\big) + e\big(G(\bar{x}), \bar{z}\big), \end{aligned}$$

where the last inequality is due to the definition of $e\big(G(\bar{x}), \bar{z}\big)$. Since $w_0 \in G(x_0)$, it holds that

$$d\big(w_0, G(\bar{x})\big) \leqslant d^{\mathcal{H}}\big(G(x_0), G(\bar{x})\big) \leqslant L\,\|x_0 - \bar{x}\|\,.$$

Consequently,

$$\|y_0 - \bar{y}\| \leqslant \|y - (\bar{y} + \bar{z})\| + L'\,\|x_0 - \bar{x}\| + e\big(G(\bar{x}), \bar{z}\big) < s' + L'r' + \eta < s.$$

From the assumption of metric regularity, we deduce

$$d\big(x_0, \Phi^{-1}(y_0)\big) \leqslant \kappa d\big(y_0, \Phi(x_0)\big) = \kappa d\big(y - w_0, \Phi(x)\big) < \kappa R. \tag{3.7}$$

On the other hand, we also have $(\bar{x}, y_0) \in \mathcal{V}$, which yields

$$d\big(\bar{x}, \Phi^{-1}(y_0)\big) \leqslant \kappa d\big(y_0, \Phi(\bar{x})\big) \leqslant \kappa\,\|y_0 - \bar{y}\| < \kappa\,(s' + L'r' + \eta)\,.$$

Hence, the quantity $d\big(x_0, \Phi^{-1}(y_0)\big)$ can be estimated as follows

$$d\big(x_0, \Phi^{-1}(y_0)\big) \leqslant \|x_0 - \bar{x}\| + d\big(\bar{x}, \Phi^{-1}(y_0)\big) < (1 + \kappa L')\,r' + \kappa s' + \kappa\eta. \tag{3.8}$$

Thanks to (3.7) and (3.8), there is a point $x_1 \in \Phi^{-1}(y_0)$ satisfying

$$\|x_0 - x_1\| < \min\big\{\kappa R, (1 + \kappa L')\,r' + \kappa s' + \kappa\eta\big\} = \beta. \tag{3.9}$$

To see that x_1 is not outside the ball $\mathbb{B}(\bar{x}, r)$, we use (3.9) and invoke (3.6)

$$\begin{aligned}\|x_1 - \bar{x}\| &\leqslant \|x_1 - x_0\| + \|x_0 - \bar{x}\| < (1 + \kappa L)\,r' + \kappa s' + \kappa\eta + r' \\ &= (2 + \kappa L)r' + \kappa s' + \kappa\eta < r.\end{aligned}$$

Remind $w_0 \in G(x_0)$. By virtue of Lipschitz continuity for G, one has

$$d\big(w_0, G(x_1)\big) \leqslant d^{\mathcal{H}}\big(G(x_0), G(x_1)\big) \leqslant L\,\|x_0 - x_1\| < L'\beta.$$

Thus, there exists one element $w_1 \in G(x_1)$ such that $\|w_0 - w_1\| < L'\beta$.

Passing to the inductive step, let's assume $x_0 \in X, \ldots, x_k \in X$ and $w_0 \in G(x_0), \ldots, w_k \in G(x_k)$ to be given. Additionally, suppose that

- $x_{i+1} \in \Phi^{-1}(y_i)$, with $y_i = y - w_i$, $i = 0, \ldots, k-1$;
- $\|x_i - x_{i+1}\| < (\kappa' L')^j\,\beta$, for each $i \leqslant k-1$;
- $\|w_i - w_{i+1}\| < L'\,(\kappa' L')^j\,\beta$, $i = 0, \ldots, k-1$.

If $x_k = x_{k-1}$, we simply define $x_{k+1} = x_k$. Otherwise, the next iteration x_{k+1} is obtained via the following procedure. First, according to the induction hypothesis, triangle inequality yields

$$\begin{aligned}\|x_k - \bar{x}\| &\leqslant \sum_{j=0}^{k-1} \|x_{j+1} - x_j\| + \|x_0 - \bar{x}\| < \sum_{j=0}^{k-1} (\kappa' L')^j \beta + r' \\ &\leqslant \frac{1}{1-\kappa' L'} \big[(1+\kappa' L')\, r' + \kappa s' + \kappa \eta \big] + r' \\ &= \frac{2}{1-\kappa' L'} r' + \frac{\kappa}{1-\kappa' L'} s' + \frac{\kappa}{1-\kappa' L'} \eta < r.\end{aligned}$$

This means $x_k \in \mathbb{B}(\bar{x}, r)$. Denoting $y_k = y - w_k$, we get

$$\begin{aligned}\|y_k - \bar{y}\| &\leqslant \sum_{j=0}^{k-1} \|y_{j+1} - y_j\| + \|y_0 - \bar{y}\| = \sum_{j=0}^{k-1} \|w_{j+1} - w_j\| + \|y_0 - \bar{y}\| \\ &< \sum_{j=0}^{k-1} L' \left(\kappa' L'\right)^j \beta + s' + L r' + \eta \\ &\leqslant \frac{L'}{1-\kappa' L'} \big[(1+\kappa L')\, r' + \kappa s' + \kappa \eta \big] + s' + L' r' + \eta \\ &= \frac{2L'}{1-\kappa' L'} r' + \frac{1}{1-\kappa' L'} s' + \frac{1}{1-\kappa' L'} \eta < s.\end{aligned}$$

Based on the fact $\kappa \in \operatorname{Regmod}\big(\Phi, \mathcal{V}\big)$, it follows that

$$d\big(x_k, \Phi^{-1}(y_k)\big) \leqslant \kappa d\big(y_k, \Phi(x_k)\big) \leqslant \kappa \|y_k - y_{k-1}\| = \kappa \|w_k - w_{k-1}\|.$$

Taking into account $\|w_{k-1} - w_k\| < L' \left(\kappa' L'\right)^{k-1} \beta$, the set $\Phi^{-1}(y_k)$ contains an element x_{k+1} satisfying $\|x_k - x_{k+1}\| < \kappa L' \left(\kappa' L'\right)^{k-1} \beta \leqslant \left(\kappa' L'\right)^k \beta$. At last, from the Lipschitz continuity hypothesis of G, the evaluations

$$d\big(w_k, G(x_{k+1})\big) \leqslant d^{\mathcal{H}}\big(G(x_k), G(x_{k+1})\big) \leqslant L \|x_k - x_{k+1}\| < L'(\kappa' L')^k \beta$$

permit us to select $w_{k+1} \in G(x_{k+1})$ with $\|w_k - w_{k+1}\| < L'(\kappa' L')^k \beta$.

Thanks to the construction above, both (x_k) and (w_k) are Cauchy sequences, so there exist the limits $x^* = \lim_{k\to\infty} x_k$ and $w^* = \lim_{k\to\infty} w_k$. Recalling $y - w_k = y_k \in \Phi(x_k)$ and $w_k \in G(x_k)$, we obtain $y - w^* \in \Phi(x^*)$ and $w^* \in G(x^*)$ after letting $k \to \infty$. Thus, $y \in \Phi(x^*) + G(x^*) = \Psi(x^*)$. By (3.9), the inequalities

$$d\big(x, \Psi^{-1}(y)\big) \leqslant \|x - x^*\| = \|x_0 - x^*\| \leqslant \sum_{k \geqslant 0} \|x_k - x_{k+1}\| \leqslant \sum_{k \geqslant 0} (\kappa' L')^k \beta$$

tell us $d\big(x, \Psi^{-1}(y)\big) \leqslant \frac{1}{1-\kappa' L'}\kappa R$. Observe that κ' and L' are taken independently from x and y, whereas the quantity R can be arbitrarily close to $d\big(y, \Psi(x)\big)$. Thus, we deduce

$$d\big(x, \Psi^{-1}(y)\big) \leqslant \frac{1}{1-\kappa L}\kappa d\big(y, \Psi(x)\big) = \tau d\big(y, \Psi(x)\big),$$

and the proof is thereby completed. □

To end up the current section, we consider a continuation version of Theorem 3.2. More precisely, we have the following theorem.

Theorem 3.4 (semi-local stability with additive perturbation)**.** *Consider a closed multifunction Φ between two Banach spaces X and Y, and let $r > 0$, $s > 0$ and $\kappa > 0$ be such that $\kappa \in \operatorname{Regmod}\big(\Phi, V(\Phi, \bar{x}, r, s)\big)$ with $\bar{x} \in \operatorname{Dom}(\Phi)$. Let L, r' and s' be some positive constants fulfilling*

$$L\kappa < 1, r' + \frac{\kappa}{1-L\kappa}s' < r, s' \leqslant s.$$

Consider a set-valued map $G : X \rightrightarrows Y$ so that both G and $\Psi = \Phi + G$ are closed mappings. If $L \in \operatorname{Lipmod}(G, \mathbb{B}(\bar{x}, r))$ then $\tau \in \operatorname{Regmod}\big(\Phi, V(\Psi, \bar{x}, r', s')\big)$ with $\tau = (1 - L\kappa)^{-1}\kappa$.

Obviously, by taking $r' = \frac{r}{4}$ and $s' = \min\left\{s, \frac{r}{5\tau}\right\}$ in Theorem 3.4, we recover Theorem 3.2 immediately.

Proof. We are able to select $\alpha > 0$ such that

$$\lambda = (L+\alpha)\kappa < 1, r' + \frac{\kappa}{1-(L+\alpha)\kappa}s' < r.$$

Fix a pair $(x, y) \in V(\Psi, \bar{x}, r', s')$ with $y \notin \Psi(x)$. Pick a constant $R > 0$ satisfying $d\big(y, \Psi(x)\big) < R < s'$. The strategy is analogous as in proof of Theorem 3.3, that generates a sequence $x_k \to x^*$ and x^* verifies the constraints below

$$y \in \Psi(x^*), \|x - x^*\| \leqslant \frac{\kappa}{1-(L+\alpha)\kappa}R. \tag{3.10}$$

Indeed, let $x_0 = x$, and choose some $w_0 \in G(x_0)$ with $d\big(y, w_0 + \Phi(x_0)\big) < R$. By setting $y_0 = y - w_0$, inclusion $(x_0, y_0) \in V\big(\Phi, \bar{x}, r, s\big)$ is valid as well. Thus,

$$d\big(x_0, \Phi^{-1}(y_0)\big) \leqslant \kappa d\big(y_0, \Phi(x_0)\big) < \kappa R,$$

which shows that there is a point, says x_1, such that $\|x_0 - x_1\| < \kappa R$. Thanks to the triangle inequality

$$\|x_1 - \bar{x}\| \leqslant \|x_1 - x_0\| + \|x_0 - \bar{x}\| < \kappa R + r' < \frac{\kappa}{1-\lambda} s' + r' < r,$$

so x_1 lies inside the ball $\mathbb{B}(\bar{x}, r)$.

Proceeding by induction, let's suppose that $x_1, \ldots, x_k \in \mathbb{B}(\bar{x}, r)$ are given points for some $k \geqslant 1$. Moreover, according to the arguments above, it is possible to assume that the following conditions hold

- $\|x_j - x_{j+1}\| < \lambda^j \kappa R$; $j = 0, \ldots, k-1$;
- $x_{j+1} \in \Phi^{-1}(y_j)$, where $y_j = y - w_j$ and $w_j \in G(x_j)$, for $j \leqslant k-1$.

If $x_k = x_{k-1}$, we simply define $x_{k+1} = x_k$. Otherwise, by using the Lipschitz continuity property for G, one has

$$d\big(w_{k-1}, G(x_k)\big) \leqslant d^{\mathcal{H}}\big(G(x_{k-1}), G(x_k)\big) \leqslant L \|x_{k-1} - x_k\|.$$

Recalling $x_{k-1} \neq x_k$, there exists an element $w_k \in G(x_k)$ agreeing with $\|w_{k-1} - w_k\| < (L+\alpha)\|x_{k-1} - x_k\|$. Setting now $y_k = y - w_k$, we find

$$\begin{aligned} d\big(y_k, \Phi(x_k)\big) &\leqslant \|y_k - y_{k-1}\| = \|w_k - w_{k-1}\| < (L+\alpha)\|x_{k-1} - x_k\| \\ &< (L+\alpha)\lambda^{k-1}\kappa R = \lambda^k R < s. \end{aligned}$$

This implies $(x_k, y_k) \in V\big(\Phi, \bar{x}, r, s\big)$. Hence,

$$d\big(x_k, \Phi^{-1}(y_k)\big) \leqslant \kappa d\big(y_k, \Phi(x_k)\big) < \kappa(L+\alpha)\|x_{k-1} - x_k\| = \lambda \|x_{k-1} - x_k\|.$$

Consequently, the set $\Phi^{-1}(y_k)$ must contain some point, called by x_{k+1}, which satisfies the estimate below

$$\|x_k - x_{k+1}\| < \lambda \|x_{k-1} - x_k\| < \lambda^k \kappa R.$$

Moreover, invoking triangle inequality many times, we deduce

$$\|x_{k+1} - \bar{x}\| \leqslant \sum_{j=0}^{k} \|x_k - x_{j+1}\| + \|x_0 - \bar{x}\| < \sum_{j=0}^{k} \lambda^k \kappa R + r' < \frac{\kappa}{1-\lambda} s' + r' < r.$$

That is, x_{k+1} belongs to $\mathbb{B}(\bar{x}, r)$. By inductive principle, the sequence (x_k) is well-defined, and indeed convergent. The rest of proof is analogous to the one of Theorem 3.3. □

3.2 Local Convergence Analysis

This section is left to the local convergence results for Josephy-Newton framework under conditions of type Kantorovich and Smale. First of all, we deal with the Kantorovich-type theorem.

Theorem 3.5 (Kantorovich-type version of local analysis)**.** *Consider problem* (3.1) *where $f : X \longrightarrow Y$ is C^2 in an open convex subset U of X. Let $x^* \in U$ be a solution of* (3.1) *and $y^* = Df(x^*)(x^*) - f(x^*)$. Suppose that $\tau \in \mathrm{Regmod}(\Phi, V_{r,s})$, where $\Phi(\cdot) := Df(x^*)(\cdot) + F(\cdot)$ and $V_{r,s} = \mathbb{B}(x^*, r) \times \mathbb{B}(y^*, s)$, with $\bar{\mathbb{B}}(x^*, r) \subset U$. Define a few quantities*

$$K(\tau, x^*, r) := \tau \sup_{\|z-x^*\| \leqslant r} \left\| D^2 f(z) \right\|, \ \textit{and} \quad \varepsilon = \min\left\{r, s, \tau s\right\}. \tag{3.11}$$

If $2K(\tau, x^*, r)r < 1$, then for each $x \in \mathbb{B}(x^*, \varepsilon)$, the algorithm (3.2) *generates a sequence (x_k) initiating at $x_0 = x$ and converging quadratically to x^**

$$\|x_{k+1} - x^*\| \leqslant \frac{1}{2r} \|x_k - x^*\|^2, k = 0, 1, 2, \dots \tag{3.12}$$

Proof. For shortness, let us denote $K^* = K(\tau, x^*, r)$. By using the mean value theorem for f, we achieve the following estimations

$$\|Df(x) - Df(x')\| \leqslant \tau^{-1} K^* \|x - x'\| \tag{3.13}$$

and

$$\|f(x) - f(x') - Df(x')(x - x')\| \leqslant \frac{1}{2}\tau^{-1} K^* \|x - x'\|^2 \tag{3.14}$$

whenever x and x' are in $\mathbb{B}(x^*, \varepsilon)$. We define $L = \frac{K^* r}{\tau} \leqslant \frac{1}{2\tau}$, $\bar{\tau} = \frac{\tau}{1-\tau L}$ and $\nu = \frac{\varepsilon}{4\tau}$. Then, it is possible to check that

$$\nu\bar{\tau} < r/2, \nu(\bar{\tau} + \tau) < r, \nu(1 + c\bar{\tau}) < s,$$

for $c = \max\{1, \tau^{-1}\}$. Let's take a parameter $\mu > 0$ for which the following relations are fulfilled

$$2\mu + \nu\bar{\tau} < r/2, \nu(\bar{\tau} + \tau) < r, 2c\mu + \nu(1 + c\bar{\tau}) < s.$$

Fixing now $x_0 \in \mathbb{B}(x^*, \varepsilon)$. We set $z_0 = Df(x_0)(x^*) - f(x^*)$ and $\Phi_0(\cdot) = Df(x_0)(\cdot) + F(\cdot)$, then $\Phi_0 = \Phi + g_0$, where $g_0 = Df(x_0) - Df(x^*)$ is a linear perturbation. Invoking (3.13), we conclude $\|g_0\| \leqslant L$. So, by applying Theorem 3.1 with the data $a = r$, $b = s$ and $\delta = r$, one has $\bar{\tau} \in \mathrm{Regmod}(\Phi_0, \mathcal{V}_0)$ for the neighborhood $\mathcal{V}_0 = \mathbb{B}(x^*, \mu) \times \mathbb{B}(z_0, \nu)$. Put $y_0 = f(x_0) - Df(x_0)(x_0)$, we obtain

$$\begin{aligned}\|y_0 - z_0\| &= \|f(x^*) - f(x_0) - Df(x_0)(x^* - x_0)\| \\ &\leqslant \frac{1}{2}\tau^{-1}K^* \|x^* - x_0\|^2 < \frac{1}{2}\tau^{-1}K^* r\varepsilon \leqslant \nu\end{aligned}$$

after including (3.14). This shows that $(x^*, y_0) \in \mathcal{V}_0$, and consequently, one gets

$$d\big(x^*, \Phi_0^{-1}(y_0)\big) \leqslant \bar{\tau} d\big(y_0, \Phi_0(x^*)\big) \leqslant \bar{\tau} \|y_0 - z_0\| \leqslant \frac{K^*}{2(1-\tau L)} \|x^* - x_0\|^2 .$$

Taking into account $2rK^* < 1$ and $\tau L \leqslant 1/2$, we can select an element x_1 in the set $\Phi_0^{-1}(y_0)$ such that $\|x^* - x_1\| < \frac{1}{2r} \|x^* - x_0\|^2$. The inclusion $x_1 \in \Phi_0^{-1}(y_0)$ can be rewritten as $f(x_0) - Df(x_0)(x_0) \in Df(x_0)(x_1) + F(x_1)$, which implies that x_1 is generated by the Josephy-Newton scheme (3.2). Furthermore, since $\|x^* - x_0\| < \varepsilon \leqslant r$, we deduce $\|x^* - x_1\| < \|x^* - x_0\|$. Therefore, using x_1 as a new starting point, we obtain x_2 by a same way.

Repeating these arguments, we have a sequence (x_k) determined through (3.2) satisfying the recurrence (3.12). Thanks to the fact $\|x^* - x_0\| < r$, (3.12) gives us $\|x^* - x_k\| \leqslant \left(\frac{1}{2}\right)^{2^k - 1} \|x^* - x_0\|$, and hence, the quadratic convergence follows. □

Remark 3.6. A similar result to Theorem 3.5 was obtained by A. Dontchev in [21, Theorem 1]. It is possible to see that the assumptions here are slightly different from the ones used in [21, Theorem 1] and the conclusion of Theorem 3.5 is more precise in the sense that it gives an explicit region for starting points.

Theorem 3.5 requires the knowledge about a solution x^* to problem (3.1) as well as the Lipschitz continuity of Df around x^*. When f is analytic, one can extend the classical γ-theory in order to attain a local behavior of convergence for the scheme (3.2). The next statement is in this sense.

Theorem 3.7 (γ-type theorem for GE)**.** *Keep in mind the assumption that f is analytic in an open convex set $U \subset X$. Let $x^* \in U$ be a solution of problem* (3.1) *and $y^* = Df(x^*)(x^*) - f(x^*)$. Suppose that $\tau \in \operatorname{Regmod}\big(\Phi, \mathcal{V}\big)$, for some $\tau > 0$ and $\Phi(\cdot) = Df(x^*)(\cdot) + F(\cdot)$, $\mathcal{V} = \mathbb{B}(x^*, r) \times \mathbb{B}(y^*, s)$. Define*

$$\gamma = \gamma(\tau, f, x^*) = \sup_{k \geqslant 2} \left\{ \left[\tau \left\| \frac{D^k f(x^*)}{k!} \right\| \right]^{\frac{1}{k-1}} \right\}.$$

Let $\psi(t) = 2t^2 - 4t + 1$ and let $\rho = \frac{3-\sqrt{7}}{2} \approx 0.17712\ldots$ be the smallest root of the *equation* $2t - \psi(t) = 0$. *Pick a number* $\varepsilon > 0$ *such that* $\bar{\mathbb{B}}(x^*, \varepsilon) \subset U$ *and*

$$\varepsilon < \min\left\{\frac{1}{2\theta}r, \frac{1}{1+\theta}r', \frac{\tau}{1+\theta}r'\right\}; \quad \textit{with } \theta = \frac{(1-\rho)^2}{\psi(\rho)} > 1.$$

If $x_0 \in \bar{\mathbb{B}}(x^*, \varepsilon)$ *and* $\|x_0 - x^*\| \gamma < \rho$, *then there exists a sequence* (x_k) *produced by* (3.2) *which obeys the following recurrent relation*

$$\|x_k - x^*\| \leqslant \frac{\gamma}{\psi(\rho)} \|x_{k-1} - x^*\|^2, k = 1, 2, \dots \tag{3.15}$$

In other words, x_k *converges Q-quadratically to* x^*.

Proof. Observe that if (3.15) is fulfilled, then we can prove that

$$\|x_k - x^*\| \leqslant \left(\frac{\gamma}{\psi(\rho)} \|x_0 - x^*\|\right)^{2^k - 1} \|x_0 - x^*\| \leqslant \left(\frac{\rho}{\psi(\rho)}\right)^{2^k - 1} \varepsilon.$$

Thus, under the hypotheses of Theorem 3.7, the convergence is straightforward.

Let's go back the main proof. We shall need some auxiliary estimations

$$\|Df(z) - Df(x^*)\| \leqslant \tau^{-1}\gamma\frac{2 - \gamma\|z - x^*\|}{(1 - \gamma\|z - x^*\|)^2}\|z - x^*\|, \tag{3.16a}$$

$$\|f(x^*) - f(z) - Df(z)(x^* - z)\| \leqslant \tau^{-1}\frac{\gamma}{(1 - \gamma\|z - x^*\|)^2}\|z - x^*\|^2, \tag{3.16b}$$

for $z \in U \cap \mathbb{B}(x^*, \gamma^{-1}\rho)$. Indeed, by using $\left\|\frac{D^j f(x^*)}{j!}\right\| \leqslant \tau^{-1}\gamma^{j-1}$, it is possible to verify that $\limsup_{j\to\infty} \left\|\frac{D^j f(x^*)}{j!}\right\|^{1/j} \leqslant \gamma$. Thus, when $z \in U \cap \mathbb{B}(x^*, \gamma^{-1}\rho)$, the expression $f(z) = \sum_{j\geqslant 0} \frac{D^j f(x^*)}{j!}(z - x^*)^j$ holds. As a result, we find

$$Df(z) - Df(x^*) = \sum_{j\geqslant 2} j\left(\frac{D^j f(x^*)}{j!}\right)(z - x^*)^{j-1}, \tag{3.17a}$$

$$f(x^*) - f(z) - Df(z)(x^* - z) = \sum_{j\geqslant 2}(j-1)\left(\frac{D^j f(x^*)}{j!}\right)(z - x^*)^j, \tag{3.17b}$$

after differentiating with respect to z. Since $\left\|\frac{D^j f(x^*)}{j!}\right\| \leqslant \tau^{-1}\gamma^{j-1}$, (3.16a) and (3.16b) are induced from (3.17a), (3.17b) and the expansions below

$$\sum_{j\geqslant 1}(j+1)q^j = \frac{1}{(1-q)^2} - 1, \sum_{j\geqslant 1} jq^j = \frac{q}{(1-q)^2}; \quad |q| < 1.$$

Now let x_0 be in assertion of Theorem 3.7. We follow the same idea as the previous theorem. We use again the notations $\Phi_0(\cdot) = Df(x_0)(\cdot) + F(\cdot)$, $g_0 = Df(x_0) - Df(x^*)$, $z_0 = Df(x_0)(x^*) - f(x^*) \in \Phi_0(x^*)$ and $y_0 = Df(x_0)(x_0) - f(x_0)$. Set $L = \tau^{-1}\frac{\rho(2-\rho)}{(1-\rho)^2} < \tau^{-1}$, $\bar{\tau} = \frac{\tau}{1-L\tau}$, $c = \max\{1, \tau^{-1}\}$ and pick $\nu = \tau^{-1}\varepsilon > 0$. Then, by a few simple computations, the following inequalities

$$\nu\bar{\tau} < \frac{r}{2}, \nu(\bar{\tau} + \tau) < r, \nu(1 + c\bar{\tau}) < r'$$

are concomitantly valid. This permits us to select $\mu > 0$ satisfying

$$2\mu + \nu\bar{\tau} < r/2, \nu(\bar{\tau} + \tau) < r, 2c\mu + \nu(1 + c\bar{\tau}) < s.$$

In view of (3.16a), one has $\|g_0\| \leqslant L$, so Theorem 3.1 gives us $\bar{\tau} \in \operatorname{Regmod}(\Phi_0, \mathcal{V}_0)$. Here, $\mathcal{V}_0$ indicates the neighborhood $\mathbb{B}(x^*, \mu) \times \mathbb{B}(z_0, \nu)$. Denoting $\sigma_0 = \|x_0 - x^*\|$, we infer from (3.16b) that

$$\begin{aligned} \|y_0 - z_0\| &= \|f(x^*) - f(x_0) - Df(x_0)(x^* - x_0)\| \\ &\leqslant \tau^{-1}\frac{\gamma}{(1-\gamma\sigma_0)^2}\|x_0 - x^*\|^2 < \tau^{-1}\frac{\rho}{(1-\rho)^2}\varepsilon < \nu. \end{aligned}$$

In other words, the pair (x^*, y_0) belongs to $\mathcal{V}_0$. As a result, we obtain

$$\begin{aligned} d\big(x^*, \Phi_0^{-1}(y_0)\big) &\leqslant \bar{\tau} d\big(y_0, \Phi_0(x^*)\big) \leqslant \frac{\tau}{1-\tau L}\|y_0 - z_0\| \\ &\leqslant \frac{1}{1-\tau L}\frac{\gamma}{(1-\gamma\sigma_0)^2}\|x_0 - x^*\|^2 < \frac{\gamma}{\psi(\rho)}\|x_0 - x^*\|^2. \end{aligned}$$

Therefore, $\Phi_0^{-1}(y_0)$ possesses an element x_1 such that $\|x^* - x_1\| < \frac{\gamma}{\psi(\rho)}\|x_0 - x^*\|^2$. To continue, let's notice that $\frac{\gamma}{\psi(\rho)}\|x_0 - x^*\| < \frac{\rho}{(1-\rho)^2} < 1$, which implies $x_1 \in \mathbb{B}(x^*, \sigma_0)$. Thus, the preceding process can be reiterated by starting at the new point x_1 instead of x_0. Consequently, there exists a sequence (x_k) being completely defined through Josephy-Newton scheme (3.2) for which the recurrence relation (3.15) is valid at all. According to (3.15), the quadratic convergence for (x_k) follows. □

Remark 3.8. Let us notice that both Theorem 3.5 and 3.7 used the informations around the solution x^*. Generally, it is not easy to localize such a solution and to have information about the local behavior of the sum $f + F$ around x^*. So, it seems to be more useful to concentrate the assumptions only on data around a chosen starting point. This is the goal of the next section.

3.3 Extensions of Kantorovich's and α-Smale's Theorems

The current section is devoted to present some convergence results which involve only the informations imposed on the starting data. The first one is a Kantorovich-type theorem for Josephy-Newton framework of (3.2).

Theorem 3.9 (Kantorovich-type theorem)**.** *Let f and F be similar as in Theorem 3.5. For $\tau > 0$, $\varepsilon > 0$ and $z \in U$ with $\bar{\mathbb{B}}(z,\varepsilon) \subset U$ we define*

$$\beta(\tau, z) := \tau d\big(0, f(z) + F(z)\big), \quad K(\tau, z, \varepsilon) := \tau \sup_{\|z'-z\| \leqslant \varepsilon} \left\|D^2 f(z')\right\|.$$

Let $x \in U$ and $\alpha \in (0,1]$ be given. Suppose that the following conditions are fulfilled:

(i) the mapping $\Phi(\cdot) = Df(x)(\cdot) + F(\cdot)$ is metrically regular on the set $\mathcal{V} = V(\Phi, x, 4r, s)$ with a modulus $\tau > \mathbf{Reg}_{\mathcal{V}}(\Phi)$, and $\bar{\mathbb{B}}(x,r) \subset U$;

(ii) $d\big(0, f(x) + F(x)\big) < s$;

(iii) $2\beta(\tau,x)K(\tau,x,r) \leqslant \alpha$;

(iv) $2\eta\beta(\tau,x) \leqslant r$, with $\eta = \frac{1}{1+\sqrt{1-\alpha}}$.

Then, problem (3.1) admits a solution x^* such that $\|x - x^*\| \leqslant 2\eta\beta(\tau,x)$. Moreover, starting at $x_0 = x$, algorithm (3.2) generates a sequence $x_k \to x^*$ satisfying the next statement: if $\alpha < 1$, then one has

$$\|x_k - x^*\| \leqslant \frac{4\sqrt{1-\alpha}}{\alpha}\frac{\theta^{2^k}}{1-\theta^{2^k}}\beta(\tau,x), \theta = \frac{1-\sqrt{1-\alpha}}{1+\sqrt{1-\alpha}}, \tag{3.18}$$

while in the case $\alpha = 1$ it holds that

$$\|x_k - x^*\| \leqslant 2^{-k+1}\beta(\tau,x). \tag{3.19}$$

Proof. Let x satisfy the hypotheses of Theorem 3.9. For simplicity, we denote $K = K(\tau,x,r)$, $\beta = \beta(\tau,x)$. If $\beta = 0$, then there is nothing to prove. Skipping this trivial case, we consider $\beta > 0$. The proof will be subdivided into several steps.

- Generating a majorizing sequence (t_k).

Let $\omega(t) = \frac{\alpha}{4\beta}t^2 - t + \beta$ be a quadratic polynomial accepting $t^* = \frac{2}{1+\sqrt{1-\alpha}}\beta$ as the smallest real root. Following the work [40], the Newton method applied to equation $\omega(t) = 0$ with $t_0 = 0$ induces a strictly increasing sequence by relation

$t_{k+1} = t_k - \omega'(t_k)^{-1}\omega(t_k)$. Furthermore, when $\alpha < 1$, the error bound

$$\begin{cases} t^* - t_k \leqslant \frac{4\sqrt{1-\alpha}}{\alpha}\frac{\theta^{2^k}}{1-\theta^{2^k}}(t_1 - t_0) = \frac{4\sqrt{1-\alpha}}{\alpha}\frac{\theta^{2^k}}{1-\theta^{2^k}}\beta, \\ \frac{2(t_{k+1}-t_k)}{1+\sqrt{1+4\theta^{2^k}\left(1+\theta^{2^k}\right)^{-2}}} \leqslant t^* - t_k \leqslant \theta^{2^{k-1}}(t_k - t_{k-1}) \end{cases} \tag{3.20}$$

is valid. Otherwise, if $\alpha = 1$, then (3.20) is replaced by

$$\begin{cases} t^* - t_k \leqslant 2^{-k+1}(t_1 - t_0) = 2^{-k+1}\beta, \\ 2\left(\sqrt{2}-1\right)(t_{k+1} - t_k) \leqslant t^* - t_k \leqslant t_k - t_{k-1}. \end{cases} \tag{3.21}$$

Particularly, we can prove by induction that

$$t_{k+1} - t_k \leqslant \alpha\left(1 + \sqrt{1-\alpha}\right)^{-2}\beta;\ k = 1, 2, \ldots \tag{3.22}$$

- Constructing a sequence (x_k) such that $\|x_{k+1} - x_k\| < t_{k+1} - t_k$.

Let $x_0 = x$, $\Phi_0(\cdot) = Df(x_0)(\cdot) + F(\cdot)$, $\tau_0 = \tau$, $r_0 = r$ and $s_0 = s$. Denoting $\mathcal{V}_0 = V(\Phi_0, x_0, 4r_0, s_0)$, then there is $\bar{\tau}_0 \in \operatorname{Regmod}\big(\Phi_0, \mathcal{V}_0\big)$ with $\bar{\tau}_0 < \tau_0$. For $y_0 = Df(x_0)(x_0) - f(x_0)$, one has

$$d\big(y_0, \Phi_0(x_0)\big) = d\big(0, f(x_0) + F(x_0)\big) = d\big(0, f(x) + F(x)\big) < s,$$

which implies $(x_0, y_0) \in \mathcal{V}_0$. Invoking $\bar{\tau}_0 \in \operatorname{Regmod}\big(\Phi_0, \mathcal{V}_0\big)$, we find

$$d\big(x_0, \Phi_0^{-1}(y_0)\big) \leqslant \bar{\tau}_0 d\big(y_0, \Phi_0(x_0)\big) < \tau d\big(0, f(x) + F(x)\big) = \beta.$$

Thus, there exists $x_1 \in \Phi_0^{-1}(y_0)$ satisfying $\|x_0 - x_1\| < \beta = t_1 - t_0$. In addition, the inclusion $x_1 \in \Phi_0^{-1}(y_0)$ gives us $Df(x_0)(x_0) - f(x_0) \in \Phi_0(x_1)$, which is equivalent to $Df(x_0)(x_0) - f(x_0) \in Df(x_0)(x_1) + F(x_1)$. In other words, x_1 is obtained via the scheme (3.2).

We proceed to the inductive step. Assume that $x_1, \ldots, x_k$ are generated by the framework of (3.2) and $\|x_{j+1} - x_j\| < t_{j+1} - t_j$ for $j \leqslant k-1$. We have

$$\|x_k - x\| \leqslant \sum_{j=0}^{k-1}\|x_{j+1} - x_j\| < \sum_{j=0}^{k-1}(t_{j+1} - t_j) = t_k < t^* = 2\eta\beta \leqslant r.$$

Set $\Phi_k(\cdot) = Df(x_k)(\cdot) + F(\cdot)$, and $g_k = Df(x_k) - Df(x)$, then $\Phi_k = \Phi + g_k$. Using the mean value theorem for f, we can check that $\|g_k\| \leqslant \tau^{-1}K\|x_k - x\| < \tau^{-1}Kt_k$. Since $t_k < t^* = 2\eta\beta \leqslant r$, we get $\|g_k\| < \tau^{-1}Kt_k \leqslant \frac{1}{2}\tau^{-1}$. Define some parameters

$L_k = Kt_k$, $\tau_k = (1 - L_k\tau)^{-1}\tau$, $r_k = \frac{r}{4}$ and $s_k = \min\left\{s, \frac{4r}{5\tau_k}\right\}$. Applying either Theorem 3.2 or Theorem 3.4, the mapping Φ_k is metrically regular on the set $\mathcal{V}_k = V\big(\Phi_k, x, r_k, s_k\big)$ with modulus τ_k. Let $y_k = Df(x_k)(x_k) - f(x_k)$, we claim $(x_k, y_k) \in \mathcal{V}_k$. Indeed, it is sufficient to prove only $d\big(y_k, \Phi_k(x_k)\big) < s_k$. Recall that x_k satisfies (3.2), we deduce

$$\begin{aligned} d\big(y_k, \Phi_k(x_k)\big) &\leqslant \|y_k - [-f(x_{k-1}) + Df(x_{k-1})(x_{k-1})]\| \\ &= \|f(x_k) - f(x_{k-1}) - Df(x_{k-1})(x_k - x_{k-1})\| . \end{aligned}$$

Because of $\tau^{-1}K = \sup_{\|z-x\|\leqslant r} \|D^2 f(z)\|$, the mean value theorem applied to f yields

$$d\big(y_k, \Phi_k(x_k)\big) \leqslant \frac{1}{2}\tau^{-1}K \|x_k - x_{k-1}\|^2 < \tau^{-1}\frac{\alpha}{4\beta}(t_k - t_{k-1})^2 .$$

According to (3.22), it holds that $t_k - t_{k-1} \leqslant \beta$. Thus, from the hypothesis that $\beta = \beta(\tau, x) < \tau s$, the estimation $d\big(y_k, \Phi_k(x_k)\big) < \frac{\alpha}{4}s < s$ is evident.
Next, expanding the polynomial ω at center t_{k-1}, and exploiting the relation $t_k - t_{k-1} = -\omega'(t_{k-1})^{-1}\omega(t_{k-1})$ one gets $\omega(t_k) = \alpha\,(t_k - t_{k-1})^2/(4\beta)$. Thanks to the facts that $2K \leqslant \beta^{-1}\alpha$ and $\omega'(t_k) = \alpha t_k/(2\beta) - 1$, we obtain

$$\begin{aligned} d\big(y_k, \Phi_k(x_k)\big) &< \tau^{-1}\omega(t_k) = -\tau^{-1}\omega'(t_k)\,(t_{k+1} - t_k) \\ &\leqslant \tau^{-1}(1 - Kt_k)\,(t_{k+1} - t_k) = \tau_k^{-1}\,(t_{k+1} - t_k) . \end{aligned}$$

Remind $\beta \leqslant \frac{1+\sqrt{1-\alpha}}{2}r$. By invoking (3.22) once more, we deduce

$$d\big(y_k, \Phi_k(x_k)\big) \leqslant \tau_k^{-1}\alpha\left(1 + \sqrt{1-\alpha}\right)^{-2}\frac{1+\sqrt{1-\alpha}}{2}r < \frac{4r}{5\tau_k}.$$

As a summary, $d\big(y_k, \Phi_k(x_k)\big) < s_k$.
Let us now apply the metric regularity property for Φ_k

$$d\big(x_k, \Phi_k^{-1}(y_k)\big) \leqslant \tau_k d\big(y_k, \Phi_k(x_k)\big) < t_{k+1} - t_k.$$

Thus, it is possible to define x_{k+1} as an element in $\Phi_k^{-1}(y_k)$ such that $\|x_k - x_{k+1}\| < t_{k+1} - t_k$. The construction is thereby completed.

To finish the proof, we observe that

$$\|x_k - x_{k+n}\| \leqslant \sum_{j=0}^{n-1}\|x_{k+j} - x_{k+j+1}\| \leqslant \sum_{j=0}^{n-1}(t_{k+j+1} - t_{k+j}) = t_{k+n} - t_k. \tag{3.23}$$

Since $t_k \to t^*$, (3.23) allows us to conclude that (x_k) is a Cauchy sequence. Let $x^* = \lim_{k\to\infty} x_k$ and let $n \to \infty$ in (3.23), we obtain (3.18) and (3.19) from (3.20) and (3.21). □

Remark 3.10. A homologous result with Theorem 3.9 was established by Dontchev [21]. The assumptions and the conclusion of Theorem 3.9 are different from the ones proved in [21, Theorem 2]. Our hypotheses concern only the starting point x_0, while in [21, Theorem 2] the author requires the informations depending not only on x_0 but also x_1.

Remark 3.11. Kantorovich-type result was also presented in [67, Theorem 3.2]. The difference between Theorem 3.9 and [67, Theorem 3.2] lies essentially on the used assumptions. In fact, the involved parameters as well as the region of the metric regularity are different. For Theorem 3.9, one needs the metric regularity property of $\Phi(\cdot) = Df(x)(\cdot)+F(\cdot)$ on the set $V(\Phi, x, r, s)$, where x is the starting point. While, the authors supposed in [67, Theorem 3.2] the Lipschitz-like hypothesis for $Q_{\bar{x}}^{-1}$ (equivalently, $Q_{\bar{x}}$ is metrically regular around $(\bar{x}, \bar{y}) \in \mathrm{Gr}(f+F)$), where $Q_{\bar{x}}(\cdot) = f(\bar{x}) + Df(\cdot - \bar{x}) + F(\cdot)$. Additionally, the authors also required in [67, Theorem 3.2] a condition that $\lim_{x\to\bar{x}} d\big(\bar{y}, f(x) + F(x)\big) = 0$ (a kind of lower semicontinuity of $f + F$, which is almost unnecessary in Theorem 3.9). An illustration for comparing the applicability of two those results will be shown in the last section of this chapter.

In the same spirit to Theorem 3.9, an extension of Smale's α-theory was also investigated. Specifically, one has the following theorem.

Theorem 3.12 (α-Smale type theorem)**.** *Let's consider problem* (3.1) *where f is analytic on an open subset U of X. For $t > 0$ and $z \in U$ we define*

$$\begin{aligned}
\beta(t, z) &= td\big(0, f(z) + F(z)\big),\\
\gamma(t, f, z) &= \sup_{k\geqslant 2} \left\{ \left[t \left\| \frac{D^k f(z)}{k!} \right\| \right]^{\frac{1}{k-1}} \right\},\\
\alpha(t, f, z) &= \beta(t, z)\gamma(t, f, z).
\end{aligned}$$

Let $\psi(t) = 2t^2 - 4t + 1$ and let $\bar{\alpha} \approx 0.1307169\ldots$ be the smallest real root of the equation

$$2t - [\psi(t)]^2 = 0.$$

Assume that:

(i) $\tau > \mathbf{Reg}_{\mathcal{V}}(\Phi)$, *in which* $\Phi(\cdot) = Df(x)(\cdot) + F(\cdot)$ *and* $\mathcal{V} = V(\Phi, x, 4r, s)$,

(ii) $d\big(0, f(x) + F(x)\big) < s$,

(iii) $\bar{\eta}\beta(\tau, x) \leqslant r$, *with* $\bar{\eta} = \frac{\bar{\alpha}+1-\sqrt{\bar{\alpha}^2-6\bar{\alpha}+1}}{4\bar{\alpha}}$,

(iv) $\alpha(\tau, f, x) \leqslant \bar{\alpha}$.

Then, it has a solution x^* *of* (3.1) *such that* $\|x - x^*\| \leqslant \bar{\eta}\beta(\tau, x) \leqslant r$. *For the initial point* $x_0 = x$, *algorithm* (3.2) *induces a sequence* (x_k) *converging to* x^* *and obeying the following estimation*

$$\|x_k - x^*\| \leqslant C[\psi(\bar{\alpha})]^k \left(\frac{1}{2}\right)^{2^k-1} \beta(\tau, x); \quad C = \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^{2^j-1}. \tag{3.24}$$

Proof. Fix some point x satisfying Theorem 3.12. To simplify the notations, we write briefly $\beta = \beta(\tau, x)$, and $\gamma = \gamma(\tau, f, x)$. Throughout this proof, the estimate below will be useful

$$\left\|D^2 f(z)\right\| \leqslant \tau^{-1} \frac{2\gamma}{(1 - \gamma\|z - x\|)^3}, \quad \text{if } z \in U \text{ and } \gamma\|z - x\| < 1 - \frac{1}{\sqrt{2}}. \tag{3.25}$$

As in the proof of Theorem 3.7, on open set $U \cap \mathbb{B}\big(x, (1 - \sqrt{2}/2)\gamma^{-1}\big)$, the Taylor's expansion $f(z) = \sum_{j\geqslant 0} \frac{D^j f(x)}{j!}(z - x)^j$ holds. Since f is analytic, we are able to differentiate with respect to z term-by-term, and therefore

$$D^2 f(z) = \sum_{j\geqslant 2} j(j-1) \frac{D^j f(x)}{j!}(z - x)^{j-2}. \tag{3.26}$$

Let us note that $\left\|\frac{D^j f(x)}{j!}\right\| \leqslant \tau^{-1}\gamma^{j-1}$. Thus, (3.25) can be obtained by taking the norm in (3.26) and then using the identical relation $\sum_{j\geqslant 2} j(j-1)q^{j-2} = \frac{2}{(1-q)^3}$, which is valid for all $|q| < 1$.
Returning the main proof, we shall separate it into a few steps.

- Majorizing function.

Define $\omega(t) = \frac{1}{1-\gamma t} - 2\gamma t + \bar{\alpha} - 1$ and consider the equation $\omega(t) = 0$. Choose $t_0 = 0$ as a starting point, the classical α-theorem by Smale [12, 77] can be applied here. This guarantees the existence of a sequence $t_{k+1} = t_k - \omega'(t_k)^{-1}\omega(t_k)$ which is strictly increasing and converges to the smallest root $t^* = \frac{\bar{\alpha}+1-\sqrt{\bar{\alpha}^2-6\bar{\alpha}+1}}{4\gamma}$ of the equation $\omega(t) = 0$. To continue the process, we define some scalar sequences (β_k),

(γ_k) and (α_k) as follows

$$\beta_k = \left|\omega'(t_k)^{-1}\omega(t_k)\right|, \gamma_k = \sup_{j\geqslant 2}\left|\omega'(t_k)^{-1}\frac{\omega^{(j)}(t_k)}{j!}\right|^{1/(j-1)}, \alpha_k = \beta_k\gamma_k.$$

A simple computation yields $\beta_0 = \beta$ and $\alpha_0 = \bar{\alpha}$. According to [15, Lemme 133], we obtain by induction that

$$\beta_{k+1} \leqslant \frac{1-\alpha_k}{\psi(\alpha_k)}\alpha_k\beta_k, \alpha_{k+1} \leqslant \min\left\{\alpha_0, \frac{1}{[\psi(\alpha_k)]^2}\alpha_k^2\right\}. \tag{3.27}$$

Exploiting the recurrence (3.27), we simultaneously prove that $\alpha_k \leqslant \left(\frac{1}{2}\right)^{2^k-1}\alpha_0$ and $\beta_k \leqslant [\psi(\alpha_0)]^n \left(\frac{1}{2}\right)^{2^n-1}\beta_0$. Hence,

$$\alpha_k \leqslant \left(\frac{1}{2}\right)^{2^k-1}\bar{\alpha},\ \beta_k \leqslant [\psi(\bar{\alpha})]^k \left(\frac{1}{2}\right)^{2^k-1}\beta. \tag{3.28}$$

• Constructing the Josephy-Newton's sequence.

We shall generate by induction a sequence (x_k) which satisfies the error bounds $\|x_k - x_{k+1}\| < \beta_k$. Here, (β_k) is defined in the preceding step. To begin, let $x_0 = x$. By an analogous argument as in proof of Theorem 3.9, there exists x_1 satisfying (3.2) and $\|x_0 - x_1\| < \beta(\tau, x) = \beta = \beta_0$.

Passing to the induction part. Suppose that the iterations $x_1, \ldots, x_k$ are generated by (3.2) and, in addition, let's assume $\|x_j - x_{j+1}\| < \beta_j$ for $j = 0, \ldots, k-1$. We consider the subproblem of solving

$$0 \in f(x_k) + Df(x_k)(x - x_k) + F(x).$$

Put $\varPhi_k(\cdot) = Df(x_k)(\cdot - x_k) + F(\cdot)$ and $g_k(\cdot) = Df(x_k)(\cdot - x_k) - Df(x)(\cdot)$, then $\varPhi_k = \varPhi + g_k$. The perturbation g_k is obviously Lipschitz continuous with a modulus $L_k = \|Df(x_k) - Df(x)\|$. Observe that

$$\|x_j - x\| \leqslant \sum_{i=0}^{j-1}\|x_i - x_{i+1}\| < \sum_{i=0}^{j-1}\beta_i = \sum_{i=0}^{j-1}(t_{i+1} - t_i) = t_j, \tag{3.29}$$

where the facts $\beta_i = |t_{i+1} - t_i| = t_{i+1} - t_i$ are used. A direct computation gives us

$$t^* = \frac{\bar{\alpha} + 1 - \sqrt{\bar{\alpha}^2 - 6\bar{\alpha} + 1}}{4\gamma} < \left(1 - \frac{1}{\sqrt{2}}\right)\gamma^{-1}.$$

Thus, the monotonicity of the sequence (t_k) permits us to write $\|x_k - x\| < t_k \leqslant t^* < \left(1 - \frac{1}{\sqrt{2}}\right)\gamma^{-1}$. Taking into account (3.25), the mean value theorem yields

$$\begin{aligned} L_k &= \left\| \int_0^1 \left[D^2 f\big(x + t(x_k - x)\big)\,(x_k - x)\,\right] dt \right\| < \tau^{-1} \int_0^1 \frac{2\gamma t_k}{(1 - t\gamma t_k)^3}\, dt \\ &= \tau^{-1}\left[-1 + (1 - \gamma t_k)^{-2}\right]. \end{aligned}$$

We introduce some parameters $\tau_k = \frac{\tau}{1-\tau L_k}$, $r_k = \frac{r}{4}$, $s_k = \min\left\{s, \frac{4r}{5\tau_k}\right\}$ and denote $\mathcal{V}_k = V(\Psi_k, x, r_k, s_k)$. Then, using either Theorem 3.2 or Theorem 3.4, it follows that $\tau_k \in \operatorname{Regmod}\big(\Phi_k, \mathcal{V}_k\big)$. To continue, let's denote $y_k = -f(x_k)$, and $v_{k-1} = x_k - x_{k-1}$. Since x_k satisfies (3.2), we deduce

$$d\big(y_k, \Phi_k(x_k)\big) = d\big(y_k, F(x_k)\big) \leqslant \big\|y_k - \big[- f(x_{k-1}) - Df(x_{k-1})(v_{k-1})\big]\big\|\,.$$

Recall that f is analytic, so the expression

$$f(x_k) = f(x_{k-1}) + Df(x_{k-1})(v_{k-1}) + \int_0^1 (1-t) D^2 f(x_{k-1} + t v_{k-1})(v_{k-1})^2\, dt$$

holds. Taking into account (3.25) and (3.29), we are going to estimate the value of $d\big(y_k, \Phi_k(x_k)\big)$ as follows

$$d\big(y_k, \Phi_k(x_k)\big) \leqslant \tau^{-1} \int_0^1 (1-t) \frac{2\gamma}{\left[1 - \gamma(t_{k-1} + t\beta_{k-1})\right]^3} \beta_{k-1}^2\, dt.$$

It is possible to check that $\omega''(t) = \frac{2\gamma^2}{(1-\gamma t)^3}$ and $\omega'(t) = -\gamma \frac{\psi(\gamma t)}{(1-\gamma t)^2}$. Therefore,

$$\begin{aligned} d\big(y_k, \Phi_k(x_k)\big) &\leqslant \frac{1}{\tau\gamma} \int_0^1 (1-t)\omega''(t_{k-1} + t\beta_{k-1})\beta_{k-1}^2 dt \\ &= \frac{1}{\tau\gamma}\Big\{\omega(t_{k-1} + \beta_{k-1}) - [\omega(t_{k-1}) + \omega'(t_{k-1})\beta_{k-1}]\Big\} \\ &= \frac{1}{\tau\gamma}\Big\{\omega(t_k) - [\omega(t_{k-1}) + \omega'(t_{k-1})(t_k - t_{k-1})]\Big\} \\ &= \frac{1}{\tau\gamma}\omega(t_k) = \tau^{-1} \frac{\psi(\gamma t_k)}{(1-\gamma t_k)^2}(t_{k+1} - t_k). \end{aligned}$$

Here, the last equality is due to $t_{k+1} - t_k = -\omega'(t_k)^{-1}\omega(t_k)$. According to $\gamma t_k \leqslant \gamma t^* < 1 - \frac{1}{\sqrt{2}}$, it holds that $\psi(\gamma t_k) \leqslant (1-\gamma t_k)^2$. Otherwise, inasmuch as $\beta_k \leqslant [\psi(\bar{\alpha})]^k \left(\frac{1}{2}\right)^{2^k - 1} \beta \leqslant \frac{1}{2}\beta$ and $\beta = \tau d\big(0, f(x) + F(x)\big) < \tau s$, the relation $d\big(y_k, \Phi_k(x_k)\big) < s$ is now clear. On the other hand, remind $\tau_k = \frac{\tau}{1-\tau L_k}$ and

$L_k < \tau^{-1}[-1+(1-\gamma t_k)^{-2}]$, we find

$$\frac{1}{\tau_k} > \tau^{-1}\left[2-(1-\gamma t_k)^{-2}\right] = \tau^{-1}\frac{\psi(\gamma t_k)}{(1-\gamma t_k)^2}.$$

Consequently,

$$\begin{aligned} d\big(y_k, \Phi_k(x_k)\big) &< \frac{1}{\tau_k}(t_{k+1}-t_k) = \frac{1}{\tau_k}\beta_k \leqslant \frac{1}{2\tau_k}\beta \leqslant \frac{1}{2\tau_k}\frac{r}{\bar\eta} \\ &= \frac{2r}{\tau_k}\frac{\bar\alpha}{\bar\alpha+1-\sqrt{\bar\alpha^2-6\bar\alpha+1}} < \frac{4r}{5\tau_k}. \end{aligned}$$

In other words, the pair (x_k, y_k) belongs to $\mathcal{V}_k$. Thus,

$$d\big(x_k, \Phi_k^{-1}(y_k)\big) \leqslant \tau_k d\big(y_k, \Phi_k(x_k)\big) < \tau\frac{(1-\gamma t_k)^2}{\psi(\gamma t_k)}\left[\tau^{-1}\frac{\psi(\gamma t_k)}{(1-\gamma t_k)^2}(t_{k+1}-t_k)\right].$$

In summary, there is $x_{k+1} \in \Phi_k^{-1}(y_k)$ with $\|x_k - x_{k+1}\| < t_{k+1} - t_k = \beta_k$. The sequence (x_k) is well-defined.

To end up the proof, we show that x_k converges to some x^* obeying (3.24). In fact, the series $\sum_{k\geqslant 0}\beta_k$ is convergent, since $\beta_k = |t_{k+1}-t_k| = t_{j+1}-t_j$. By virtue of $\|x_k - x_{k+1}\| < \beta_k$, (x_k) is a convergent sequence. Letting $x^* = \lim\limits_{k\to\infty} x_k$ and taking into account the following relations

$$\|x - x^*\| = \|x_0 - x^*\| \leqslant \sum_{k\geqslant 0}\|x_k - x_{k+1}\| \leqslant \sum_{k\geqslant 0}\beta_k = \sum_{k\geqslant 0}(t_{k+1}-t_k) = t^*,$$

the inequality $\|x - x^*\| \leqslant \bar\eta\beta(\tau, x)$ is evident. Finally, to obtain (3.24), we invoke $\|x_k - x^*\| \leqslant \sum_{j\geqslant k}\|x_j - x_{j+1}\| \leqslant \sum_{j\geqslant k}\beta_j$ and then apply (3.28). □

Remark 3.13. Theoretically, we can improve slightly the value of $\bar\alpha$ in Theorem 3.12. Indeed, it is possible to take $\bar\alpha$ as any positive number $a < 1 - \frac{1}{\sqrt{2}}$ such that

$$\sup_{0\leqslant t\leqslant a}\frac{t}{[\psi(t)]^2} = q(a) < 1. \tag{3.30}$$

(Concerning Theorem 3.12, $q(\bar\alpha)$ is equal to $\frac{1}{2}$.) In his work [83], the author has developed the notion of Lipschitz continuity with L-average, and then applied to the study of Smale-type theory. Towards this development, he obtained a very good value $\bar\alpha' = 3 - 2\sqrt{2} \approx 0.17157\ldots$ Observe that if $\bar\alpha$ is replaced by $\bar\alpha'$ in Theorem 3.12, then the majorizing equation $\omega_1(t) = \frac{1}{1-\gamma t} - 2\gamma t + \bar\alpha' - 1 = 0$ will admit $\left(1 - \frac{1}{\sqrt{2}}\right)\gamma^{-1}$ as the smallest real solution.

Remark 3.14. As a particular case, problem (3.1) becomes nonlinear equation $f(x) = 0$ under restriction $F = 0$. Thus, it will be very interesting to expect either Theorem 3.9 or Theorem 3.12 recovers the corresponding classical theorem for Newton's method of solving equation. Unfortunately, this seems to be impossible. More precisely, the Kantorovich theorem cannot be recovered from Theorem 3.9 by letting $F = 0$ directly. A similar argument is also true for Theorem 3.12. Those failures are due to the fact that our involved parameters might be not the same as the classical ones.

3.4 Some Examples

Example 3.15 (An illustration for Remark 3.11)**.** The purpose of this example is to sketch a comparison mentioned in Remark 3.11. We test a simple case in one dimension including $f(x) = \frac{1}{3}x^3 - x + 1$ and $F(x) = [0, +\infty)$, $x \in \mathbb{R}$. Let's choose the reference point $x_0 = -2$ and fix $r = 0.5$, $s = 1$. By setting $\varPhi_{x_0}(u) := f'(x_0)(u) + F(u)$, $u \in \mathbb{R}$, it is easy to verify the following succession of equalities

$$\begin{aligned} d\big(u, \varPhi_{x_0}^{-1}(v)\big) &= \max\left\{0, u - \frac{v}{f'(x_0)}\right\} = \max\left\{0, u - \frac{1}{3}v\right\} \\ &= \frac{1}{3}\max\Big\{0, f'(x_0)u - v\Big\} = \frac{1}{3}d\big(v, \varPhi_{x_0}(u)\big). \end{aligned}$$

As a result, $\mathbf{Reg}_{V(\varPhi_{x_0}, x_0, 4r, s)}(\varPhi_{x_0}) = \frac{1}{3}$. Pick $\tau = 0.5 > \frac{1}{3}$, we are able to write $\beta(\tau, x_0) = \frac{1}{6}$, $K(\tau, x_0, r) \leqslant \frac{5}{6}$. Taking $\alpha = \frac{5}{18} < 1$, one has $2\eta\beta = \frac{1}{3}\frac{1}{1+\sqrt{1-\alpha}} < \frac{1}{3} < r$. These arguments show that all conditions of Theorem 3.9 are valid at initial point $x_0 = -2$.

Let us check whether the assumptions of Theorem 3.2 in [67] hold at $x_0 = -2$. Here, we have $Q_{x_0}(u) = f(x_0) + f'(x_0)(u - x_0) + \mathbb{R}_+$. The same notations M, L, r_{x_0}, r_{y_0}, r_0, δ, η and $y_0 \in Q_{x_0}(x_0)$ as in [67, Theorem 3.2] will be used. To apply [67, Theorem 3.2] at $x_0 = -2$, the system of constraints below must be fulfilled

$$\begin{cases} r_0 = \min\left\{r_{y_0}, -2Lr_{x_0}^2, \frac{r_{x_0}(1 - MLr_{x_0})}{4M}\right\}, \\ \delta \leqslant \min\left\{\frac{r_{x_0}}{4}, \frac{r_{y_0}}{11L}, 6r_0, 1\right\}, \\ (M+1)L(\eta\delta + 2r_{x_0}) \leqslant 2, \\ |y_0| < \frac{L\delta^2}{4}. \end{cases} \tag{3.31}$$

We notice that $y_0 \in f(x_0) + \mathbb{R}_+ = \left[\frac{1}{3}, +\infty\right)$, $M \geqslant \operatorname{Reg} Q_{x_0}(x_0, y_0) = \frac{1}{3}$, while the constant L should satisfy

$$L = \sup_{|u-x_0|\leqslant r_{x_0}/2} |f''(u)| = \sup_{|u-x_0|\leqslant r_{x_0}/2} |2u| = 2\,|x_0| + r_{x_0} = 4 + r_{x_0}.$$

From the second and the last inequalities in (3.31), we deduce

$$\frac{1}{3} \leqslant |y_0| < \frac{L\delta^2}{4} \leqslant \frac{L}{64} r_{x_0}^2 = \frac{r_{x_0}^2(4+r_{x_0})}{64}. \tag{3.32}$$

Otherwise, it follows from the third inequality of (3.32) that

$$(4+r_{x_0})r_{x_0} = Lr_{x_0} < \frac{1}{M+1} \leqslant \frac{3}{4}. \tag{3.33}$$

However, (3.32) and (3.33) may not be simultaneously valid. In summary, the result in [67, Theorem 3.2] seems to be not applicable at $x_0 = -2$.

Example 3.16 (Feasibility problem with Kantorovich's approach)**.** A feasibility problem associated with a C^1 function $g : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ and subset $\emptyset \neq K \subset \mathbb{R}^n$ is of the form

$$\text{find} \quad x \in \mathbb{R}^m \quad \text{such that} \quad g(x) \in K. \tag{FP}$$

By simply setting $f(x) = -g(x)$ as well as $F(x) := K$, (FP) becomes a GE $0 \in f(x) + F(x)$. The JNm applied to (FP) reads

$$g(x_k) + Dg(x_k)(u_k) \in K, x_{k+1} = x_k + u_k. \tag{JNFP}$$

Let $\Psi_x(\cdot) := -Dg(x)(\cdot) + K$. Due to the sum rule in Theorem 2.4, one obtains

$$\widehat{D}^*\Psi_x(u,w)(w^\star) = \begin{cases} -\nabla g(x)^T w^\star, & \text{if } -w^\star \in \widehat{N}_K\big(w + \nabla g(x)u\big),\\ \emptyset, & \text{if } -w^\star \notin \widehat{N}_K\big(w + \nabla g(x)u\big),\end{cases}$$

and concomittantly

$$D^*\Psi_x(u,w)(w^\star) = \begin{cases} -\nabla g(x)^T w^\star, & \text{if } -w^\star \in N_K\big(w + \nabla g(x)u\big),\\ \emptyset, & \text{if } -w^\star \notin N_K\big(w + \nabla g(x)u\big).\end{cases}$$

Here, $\nabla g(x)^T$ is the transpose of the Jacobian $\nabla g(x)$. Thanks to [59], we get

$$\operatorname{Reg}\Psi_x(u,w) = \left|D^*\Psi_x(u,w)^{-1}\right|^+, \tag{3.34}$$

$$\operatorname{Greg}\Psi_x(u) = \inf_{\alpha>0} \sup_{\substack{\|u'-u\|\leqslant\alpha,\\ w'\in\Psi_x(u')}} \left\{ \left| \widehat{D}^*\Psi_x(u',w')^{-1} \right|^+ \right\}, \tag{3.35}$$

where $\operatorname{Greg}\Psi_k(u)$ is the infimum of all moduli $\kappa > 0$ for which the mapping Ψ_k is metrically regular on some set $\big\{(v,z) : \|v-u\| \leqslant \mu, d\big(z, \Psi_k(v)\big) \leqslant \nu\big\}$.

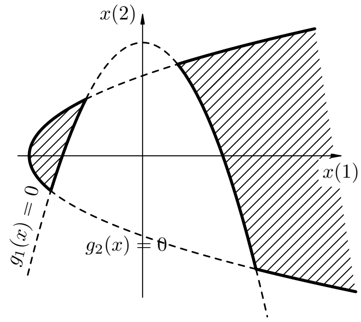
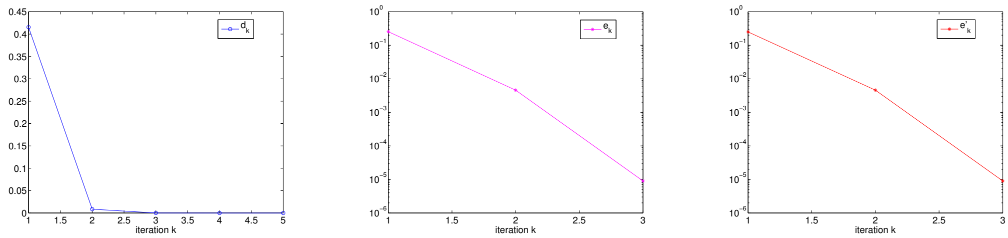


Fig. 3.2: The feasible set $\big\{g(x) \in K\big\}$ in Example 3.16

As an illustration, let us take $m = n = 2$, $K = \mathbb{R}^2_+$ and $g_1(x) = x(1)^2 + x(2) - 2$, $g_2(x) = x(1) - x(2)^2 + 2$ for $x = (x(1), x(2))^T$ in $\mathbb{R}^2$. Figure 3.2 depicts the feasible set corresponding to such g and K. We carry out the tests by applying the Kantorovich-type theorems (i.e. Theorem 3.5 and Theorem 3.9). Figures 3.3 and 3.4 describe the numerical results under several certain choices of starting point.

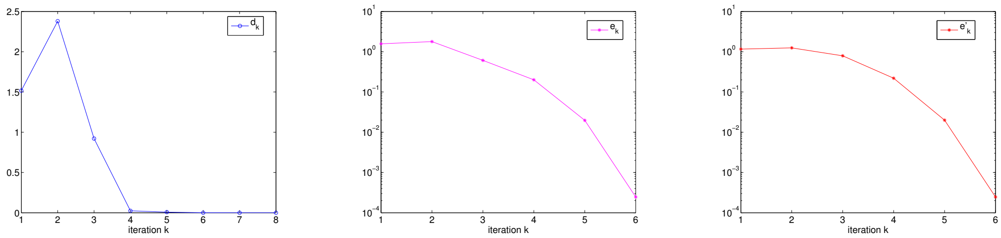


(a): $d_k = d(0, f(x_k) + F(x_k))$ (b): $e_k = \|x_{k+1} - x_k\|$ (c): $e'_k = \|x_k - x_{kmax}\|$

Fig. 3.3: Numerical test in Example 3.16: $x_0 = (-0.903, 0.77)^T$

Example 3.17 (Complementarity problem)**.** A complementarity problem corresponding to a map $h : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ and a fixed cone $\emptyset \neq C \subset \mathbb{R}^m$ can be written as follows (see [30])

$$\text{find} \quad x \in C \quad \text{such that} \quad h(x) \in C^*, x \perp h(x). \tag{CP}$$



(a): $d_k = d\big(0, f(x_k) + F(x_k)\big)$ (b): $e_k = \|x_{k+1} - x_k\|$ (c): $e'_k = \|x_k - x_{kmax}\|$

Fig. 3.4: Numerical test in Example 3.16: $x_0 = (0.039, 0.481)^T$

Here, $C^* = \big\{v : \langle v, u\rangle \geqslant 0, \text{ for all } u \in C\big\}$ is the dual cone of C, while expression $u \perp v$ means that the vector u is perpendicular to v. Under assignments $g(x) = (-x, -h(x), \langle x, h(x)\rangle)$ and $K = C \times C^* \times \{0\}$, we transform (CP) into the form of a feasibility problem described in (FP). Therefore, a same strategy as in the previous example would be applicable.

Example 3.18 (Applicability of α-type theorem)**.** In this simple example, we check $X = Y = \mathbb{R}$, $f(x) = \frac{1}{3}x^3 - x^2$ and

$$F(x) = \begin{cases} \big\{\frac{1}{2}x, x^2 - x\big\}, & \text{if } x \geqslant 0, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{3.36}$$

The graph of set-valued map $f + F$ is shown in Figure 3.5. For the data f and

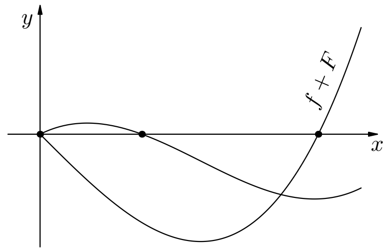


Fig. 3.5: The sum $f(\cdot) + F(\cdot)$ in Example 3.18

F in the current case, it is not difficult to compute directly the quantities $\beta(t, z)$, $\gamma(t, f, z)$ and $\alpha(t, f, z)$. In details, one has

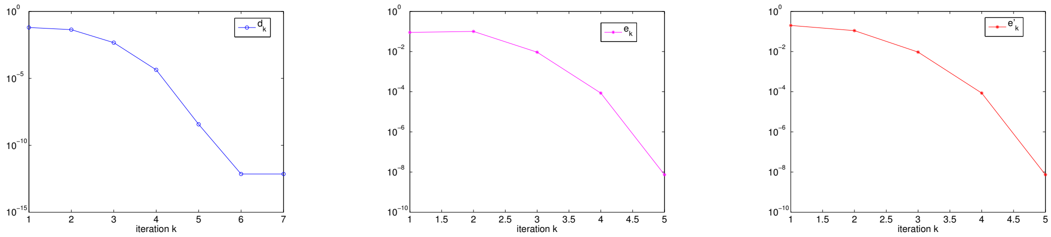
$$\beta(t, z) = t \min\left\{\left|\frac{1}{3}z^3 - z^2 + \frac{1}{2}z\right|, \left|\frac{1}{3}z^3 - z\right|\right\},$$

$$\gamma(t, f, z) = \max\left\{|t(z-1)|, \sqrt{\frac{t}{3}}\right\},$$

for $t > 0$ and $z \in \mathbb{R}$. Otherwise, by setting $\Phi_x(u) = f'(x)u + F(u)$, $u \in \mathbb{R}$, we deduce

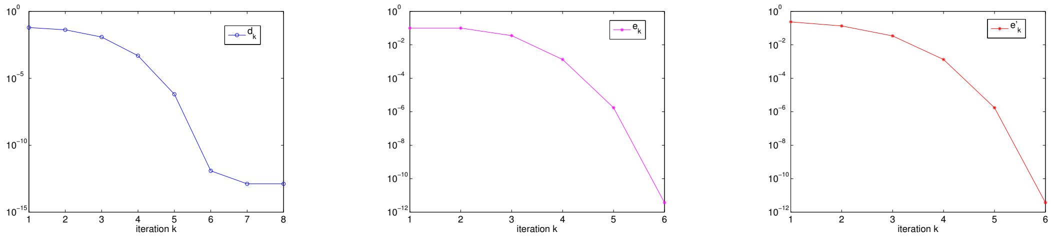
$$\begin{aligned} \operatorname{Greg} \Phi_x(\bar{u}) &= \max \left\{ \operatorname{Reg} \Phi_x(\bar{u}, \bar{v}) : \bar{v} \in \Phi_x(\bar{u}) \right\} \\ &= \max \left\{ \left| x^2 - 2x + 0.5 \right|^{-1}, \left| 2\bar{u} + x^2 - 2x - 1 \right|^{-1} \right\} \end{aligned} \tag{3.37}$$

with the convention $1/0 = +\infty$. The first equality in (3.37) is due to [60, Proposition 1.50]. Let's apply the α-type theorem (Theorem 3.12) using in turn the starting points $x_0 = 0.2$ and $x_0 = 0.4$. The tests are illustrated in the next Figures 3.6 and 3.7.



(a): $d_k = d(0, f(x_k) + F(x_k))$ (b): $e_k = |x_{k+1} - x_k|$ (c): $e'_k = |x_k - x_{kmax}|$

Fig. 3.6: Numerical test in Example 3.18: $x_0 = 0.2$, $x_{kmax} = 1.4354e - 12$



(a): $d_k = d(0, f(x_k) + F(x_k))$ (b): $e_k = |x_{k+1} - x_k|$ (c): $e'_k = |x_k - x_{kmax}|$

Fig. 3.7: Numerical test in Example 3.18: $x_0 = 0.4$, $x_{kmax} = 0.6339746\ldots$

Chapter 4

Newton-Type Method Using Set-Valued Differentiation

This chapter continues considering generalized equations between Banach spaces involving set-valued maps. However, we will approach them under another point of view. Roughly speaking, in contrast to Chapter 3, where the multivalued term is preserved during the iterative process, the current chapter suggests another kind of framework for which both single and set-valued parts are approximated. As a preparatory step, we first investigate in the forthcoming section some results dealing with metric regularity under set-valued perturbations.

4.1 Metric Regularity and Perturbed Set-valued Maps

It is well-known in the literature, e.g. [5, 26, 44] that if a given set-valued map Φ is metrically regular around a reference point $(\bar{x}, \bar{y}) \in \operatorname{Gr} \Phi$ then the sum $\Phi' = \Phi + G$, where G is a Lipschitz continuous mapping, is also metrically regular around $(\bar{x}, \bar{y} + \bar{z}) \in \operatorname{Gr} \Phi'$ $(\bar{z} \in G(\bar{x}))$ under some suitable conditions. Unfortunately, as mentioned by A.D. Ioffe [45], the additive type of perturbation are by no means representative in the category of set-valued maps. And in the same paper, Ioffe had introduced another quantity of measuring given in the definition below.

Definition 4.1. Given two Banach spaces X and Y, and let $T_1, T_2 : X \rightrightarrows Y$ be two set-valued maps. For $x \in X$ and $r > 0$, one defines

$$\sigma_{T_1,T_2}(x, r) := \sup_{\xi_2 \in T_2(x)} \inf_{\xi_1 \in T_1(x)} \sup_{\substack{d(x,u) \leqslant r, \\ \zeta_1 \in T_1(u)}} \inf_{\zeta_2 \in T_2(u)} \|\xi_2 - \xi_1 + \zeta_1 - \zeta_2\| . \tag{4.1}$$

Let's pay attention to a special case where $T_2 = T_1 + G$ is the sum of T_1 and a set-valued map $G : X \rightrightarrows Y$. According to the discussions in [45] and [63], we have $\sigma_{T_1,T_2}(x,r) \leqslant \sup_{d(x,u)\leqslant r} e\big(G(x), G(u)\big)$. Therefore, if G is Lipschitz continuous with a modulus L around a point $\bar{x} \in X$, then

$$\sigma_{T_1,T_2}(x,r) \leqslant Lr \tag{4.2}$$

whenever x is nearby $\bar{x}$ and r is small enough.

Up to the current work, let us now present the first statement related to semi-local metric regularity property.

Proposition 4.2 (stability of semi-local metric regularity)**.** *Given two closed multifunctions $\Phi : X \rightrightarrows Y$ and $\Psi : X \rightrightarrows Y$ between two Banach spaces X and Y. Let r, s, r', s', κ and μ be positively real numbers such that*

$$\kappa\mu < 1, r' + \frac{\kappa}{1-\kappa\mu}s' < r, s' < s. \tag{4.3}$$

Suppose that:

(i) Φ is metrically regular on the set

$$V_{r,s}(\Phi,\bar{x}) = \big\{(z,w) \in X \times Y : \|z-\bar{x}\| \leqslant r, d\big(w, \Phi(z)\big) \leqslant s\big\}$$

with a modulus κ;

(ii) $\sigma_{\Phi,\Psi}(x,\nu) \leqslant \mu\nu$, for all $x \in \bar{\mathbb{B}}_X(\bar{x},r)$ and $\nu \leqslant \kappa s$.

If we set $\tau = \frac{1}{1-\kappa\mu}\kappa$, then $\tau \in \operatorname{Regmod}\big(\Psi, V_{r',s'}(\Psi,\bar{x})\big)$, where

$$V_{r',s'}(\Psi,\bar{x}) = \big\{(z,w) \in X \times Y : \|z-\bar{x}\| \leqslant r', d\big(w, \Psi(z)\big) \leqslant s'\big\}.$$

Assuming $\Psi = \Phi + G$, where G is Lipschitz continuous on the ball $\mathbb{B}(\bar{x},\delta)$ for $\delta = \kappa s + r$. Then, assumption (ii) holds by (4.2). Hence, we recover Theorem 3.4 as a consequence of Proposition 4.2.

Proof. We choose some $\kappa' > \kappa$, $\mu' > \mu$ such that

$$\kappa'\mu' < 1, r' + \frac{\kappa'}{1-\kappa'\mu'}s' < r.$$

Let's fix a pair (z,w) belonging to the set $\mathcal{V}' = V_{r',s'}(\Psi,\bar{x})$. Pick a constant R

satisfying

$$d\big(w, \Psi(z)\big) < R < s, r' + \frac{\kappa'}{1-\kappa'\mu'}R < r.$$

At the first step, we set $z_0 = z$ and $\nu_0 = \kappa R < \kappa s$. By the choice of R, it is possible to take some $v_0 \in \Psi(z_0)$ with $\|w - v_0\| < R$. Under assumption (ii), one has

$$\inf_{\xi' \in \Phi(z_0)} \sup_{\substack{d(z_0,y)\leqslant \nu_0,\\ \zeta\in\Phi(y)}} \inf_{\zeta'\in\Psi(y)} \|v_0 - \xi' + \zeta - \zeta'\| \leqslant \sigma_{\Phi,\Psi}(z_0,\nu_0) \leqslant \mu\nu_0.$$

Thus, there exists a point $u_0 \in \Phi(z_0)$ having property that

$$\sup_{d(z_0,y)\leqslant\nu_0, \zeta\in\Phi(y)} \inf_{\zeta'\in\Psi(y)} \|v_0 - u_0 + \zeta - \zeta'\| < \mu'\nu_0. \tag{4.4}$$

Define $w_0 = w - v_0 + u_0$, we find $d\big(w_0, \Phi(z_0)\big) \leqslant \|w_0 - u_0\| = \|w - v_0\| < R$, which allows us to write $d\big(w_0, \Phi(z_0)\big) < s$. Particularly, this implies the inclusion $(z_0, w_0) \in V_{r,s}(\Phi, \bar{x})$. Invoking the hypothesis of metric regularity for Φ, we arrive

$$d\big(z_0, \Phi^{-1}(w_0)\big) \leqslant \kappa d\big(w_0, \Phi(z_0)\big) < \kappa R = \nu_0.$$

Let us select $z_1 \in \Phi^{-1}(w_0)$ with $\|z_0 - z_1\| < \kappa R$. Under the substitution $y = z_1$, (4.4) gives us $\inf_{\zeta'\in\Psi(z_1)} \|v_0 - u_0 + w_0 - \zeta'\| < \mu'\nu_0$. Consequently, for some $v_1 \in \Psi(z_1)$, the inequality $\|v_0 - u_0 + w_0 - v_1\| < \mu'\nu_0$ is fulfilled.

We proceed to the induction step. Suppose $z_0 = z, z_1, \ldots, z_k$ and $v_0 \in \Psi(z_0)$, $\ldots$, $v_k \in \Psi(z_k)$ are given. Towards the preceding arguments, we should require the following conditions to be valid:

- there exist $u_0 \in \Phi(z_0)$, $\ldots$, $u_{k-1} \in \Phi(z_{k-1})$ such that $z_{j+1} \in \Phi^{-1}(w_j)$ for $w_j = w - v_{j+1} + u_j$;
- $\|v_j - u_j + w_j - v_{j+1}\| < \mu'\nu_j$ with $\nu_j = (\kappa'\mu')^j\nu_0$ and $j \leqslant k-1$;
- $\|z_j - z_{j+1}\| < (\kappa'\mu')^j\kappa R$, $j = 0, \ldots, k-1$.

Define a new parameter $\nu_k = (\kappa'\mu')^k\nu_0$. Then, it is clear that $\nu_k < \nu_0 = \kappa R < \kappa s$. By virtue of (ii), the inequality $\sigma_{\Phi,\Psi}(x_k, \nu_k) \leqslant \mu\nu_k$ is straightforward. Observing $v_k \in \Psi(z_k)$, we conclude that

$$\inf_{\eta\in\Phi(z_k)} \sup_{\substack{d(z_k,y)\leqslant\nu_k,\\ \upsilon\in\Phi(y)}} \inf_{\upsilon'\in\Psi(y)} \|v_k - \eta + \upsilon - \upsilon'\| \leqslant \sigma_{\Phi,\Psi}(z_k,\nu_k) \leqslant \mu\nu_k < \mu'\nu_k.$$

This permits us to take an element $u_k \in \Phi(z_k)$ such that

$$\sup_{d(z_k,y)\leqslant\nu_k,\upsilon\in\Phi(y)} \inf_{\upsilon'\in\Psi(y)} \|v_k - u_k + \upsilon - \upsilon'\| < \mu'\nu_k. \tag{4.5}$$

Setting now $w_k = w - v_k + u_k$, we claim $(v_k, w_k) \in V_{r,s}\big(\Phi, \bar{x}\big)$. In fact, one has

$$\begin{aligned}\|z_k - \bar{x}\| &\leqslant \sum_{j=0}^{k-1} \|z_j - z_{j+1}\| + \|z_0 - \bar{x}\| < \sum_{j=0}^{k-1} \big[(\kappa'\mu')^j \kappa R\big] + r' \\ &< \frac{1}{1-\kappa'\mu'}\kappa R + r' < r.\end{aligned}$$

On the other hand, due to the choice of w_k, it is possible to estimate the distance $d\big(w_k, \Phi(z_k)\big)$ as follows

$$\begin{aligned}d\big(w_k, \Phi(z_k)\big) \leqslant \|w_k - u_k\| &= \|w - v_k\| = \|w_{k-1} + v_{k-1} - u_{k-1} - v_k\| \\ &< \mu'\nu_{k-1} = \mu'(\kappa'\mu')^{k-1}\nu_0 = (\kappa'\mu')^k R < s.\end{aligned}$$

Hence, $(v_k, w_k) \in V_{r,s}\big(\Phi, \bar{x}\big)$. Since $\kappa \in \operatorname{Regmod}\big(\Phi, V_{r,s}(\Phi, \bar{x})\big)$, we obtain

$$d\big(z_k, \Phi^{-1}(w_k)\big) \leqslant \kappa d\big(w_k, \Phi(z_k)\big) < \kappa(\kappa'\mu')^k R.$$

As a result, the set $\Phi^{-1}(w_k)$ must contain an element, written as z_{k+1}, such that $\|z_k - z_{k+1}\| < (\kappa'\mu')^k \kappa R$. Remind $\nu_k = (\kappa'\mu')^k \kappa R$, after substituting $y = z_{k+1}$ and using $w_k \in \Phi(z_{k+1})$, the estimation in (4.5) yields

$$\inf_{\upsilon'\in\Psi(z_{k+1})} \|v_k - u_k + w_k - \upsilon'\| < \mu'(\kappa'\mu')^k \kappa R.$$

From this, we can choose in the set $\Psi(z_{k+1})$ a point v_{k+1} satisfying $\|v_k - u_k + w_k - v_{k+1}\| < \mu'(\kappa'\mu')^k \kappa R$. The construction is done.

In order to obtain the necessary conclusions, we prove that the sequence (z_k) converges. Indeed, for every indices k and l with $k > l$, we have $\|z_k - z_l\| \leqslant \sum_{j=0}^{l-k-1} \|z_{k+j} - z_{k+j+1}\|$. According to the construction above, the inequality $\|z_{k+j} - z_{k+j+1}\| < (\kappa'\mu')^{k+j} \kappa R$ holds. Thus, the term $\|z_k - z_l\|$ can be dominated as follows

$$\|z_k - z_l\| < \sum_{j=0}^{l-k-1} (\kappa'\mu')^{k+j} \kappa R < (\kappa'\mu')^k \frac{1}{1-\kappa'\mu'} \kappa R. \tag{4.6}$$

Consequently, the sequence (z_k) is Cauchy, so it converges. Letting $k = 0$ and passing to the limit as $l \to \infty$ in (4.6) we get $\|z - z^*\| \leqslant \frac{1}{1-\kappa'\mu'}\kappa R$, where $z^* = \lim_{k\to\infty} z_k$. On the other side, taking into account $v_k \in \Psi(z_k)$ and $\|w - v_k\| = \|w_{k-1} + v_{k-1} - u_{k-1} - v_k\| < \mu'(\kappa'\mu')^k R$, we deduce $w = \lim_{k\to\infty} v_k \in \Psi(z^*)$. Therefore,

$$d\big(z, \Psi^{-1}(w)\big) \leqslant \|z - z^*\| \leqslant \frac{1}{1-\kappa'\mu'}\kappa R.$$

Because the quantity $\frac{1}{1-\kappa'\mu'}\kappa R$ can be arbitrarily close to $\tau d\big(w, \Psi(z)\big)$, we conclude $d\big(z, \Psi^{-1}(w)\big) \leqslant \frac{1}{1-\kappa\mu}\kappa d\big(w, \Psi(z)\big)$, and the proof is thereby completed. □

In their work [63], the authors have established some stability results related to the local metric regularity. For studying the local behavior in the next section, it is sufficient to use a weaker form. The next statement is in this sense.

Proposition 4.3. *Let $\Phi : X \rightrightarrows Y$ and $\Psi : X \rightrightarrows Y$ be two closed set-valued maps, and let $(\bar{x}, \bar{y}) \in \mathrm{Gr}\Phi$. Given some positive numbers κ, r and s such that $\kappa \in \mathrm{Regmod}\big(\Phi, \mathcal{V}\big)$ for a neighborhood $\mathcal{V} = \mathbb{B}(\bar{x}, r) \times \mathbb{B}(\bar{y}, s)$. Consider some parameters $\mu > 0$, $s' > 0$ and $\delta > 0$ with*

$$\kappa\mu < 1, \frac{\kappa}{1-\kappa\mu}s' < r, \delta + (1+\kappa\mu)s' < s. \tag{4.7}$$

Assume that there is $\bar{z} \in \Psi(\bar{x})$ for which the condition

$$\inf_{v\in\Phi(x)} \sup_{w\in\Psi(x)} \|(w - \bar{z}) - (v - \bar{y})\| \leqslant \delta \tag{4.8}$$

holds whenever $x \in \mathbb{B}_X(\bar{x}, r)$. If

$$\sigma_{\Phi,\Psi}(x, \varepsilon) \leqslant \mu\varepsilon, \quad \text{for} \quad x \in \mathbb{B}(\bar{x}, r) \quad \text{and} \quad \varepsilon < r, \tag{4.9}$$

then one has

$$d\big(\bar{x}, \Psi^{-1}(z)\big) \leqslant \frac{\kappa}{1-\kappa\mu} d\big(z, \Psi(\bar{x})\big); \quad z \in Y, \|z - \bar{z}\| < s'. \tag{4.10}$$

Proof. Let's first take $\kappa' > \kappa$ and $\mu' > \mu$ such that

$$\kappa'\mu' < 1, \frac{\kappa'}{1-\kappa'\mu'}s' < r, (1+\kappa'\mu')s' + \delta < s.$$

Fix a point $z \in \mathbb{B}(\bar{x}, s')$, our goal is that, to establish the inequality $d\big(\bar{x}, \Psi^{-1}(z)\big) \leqslant \frac{\kappa'}{1-\kappa'\mu'} d\big(z, \Psi(\bar{x})\big)$. Thus, we shall produce a suitable approximating sequence (x_k) initiating at $x_0 = \bar{x}$. Denoting $C = d\big(z, \Psi(\bar{x})\big)$, then $C \leqslant \|z - \bar{z}\| < s'$. Let C'

belong to the interval (C, s'). We select an element $w_0 \in \Psi(x_0)$ with $\|z - w_0\| < C'$ and put $\nu_0 = \kappa' C' > 0$. From (4.9), it is possible to write down

$$\inf_{v\in\Phi(x_0)} \sup_{\|x-x_0\|\leqslant\nu_0,\xi\in\Phi(x)} \inf_{\xi'\in\Psi(x)} \|w_0 - v + \xi - \xi'\| \leqslant \sigma_{\Phi,\Psi}(x_0,\nu_0) \leqslant \mu\nu_0.$$

Therefore, it has some $v_0 \in \Phi(x_0)$ which satisfies the property below

$$\sup_{\|x-x_0\|\leqslant\nu_0,\xi\in\Phi(x)} \inf_{\xi'\in\Psi(x)} \|w_0 - v_0 + \xi - \xi'\| < \mu'\nu_0. \tag{4.11}$$

Define a new point $y_0 := z + v_0 - w_0$, then one has

$$\begin{aligned}\|y_0 - \bar{y}\| = \|z - v_0 - w_0 - \bar{y}\| &\leqslant \|z - \bar{z}\| + \|\bar{z} - v_0 - w_0 - \bar{y}\| \\ &\leqslant \|z - \bar{z}\| + \|w_0 - v_0 + \xi - \xi'\| + \|-\xi + \bar{y} + \xi' - \bar{z}\|\end{aligned}$$

for any $\xi \in \Phi(x_0)$ and $\xi' \in \Psi(x_0)$. Particularly,

$$\begin{aligned}\|y_0 - \bar{y}\| \leqslant \|z - \bar{z}\| + &\sup_{\xi\in\Phi(x_0)} \inf_{\xi'\in\Psi(x_0)} \|w_0 - v_0 + \xi - \xi'\| \\ &+ \inf_{\xi\in\Phi(x_0)} \sup_{\xi'\in\Psi(x_0)} \|-\xi + \bar{y} + \xi' - \bar{z}\|.\end{aligned}$$

In view of (4.8) and (4.11), we obtain

$$\|y_0 - \bar{y}\| < s' + \mu'\nu_0 + \delta = s' + \kappa'\mu'C' + \delta < (1 + \kappa'\mu')s' + \delta < s.$$

Combining with $\|x_0 - x\| < r'$, the pair (x_0, y_0) belongs to $\mathcal{V}$. Thus, it infers from the metric regularity of Φ that

$$d\big(x_0, \Phi^{-1}(y_0)\big) \leqslant \kappa' d\big(y_0, \Phi(x_0)\big) \leqslant \kappa' \|y_0 - v_0\| = \kappa' \|z - w_0\| < \kappa' C'.$$

This ensures the existence of a point $x_1 \in \Phi^{-1}(y_0)$ such that $\|x_0 - x_1\| < \kappa' C' = \nu_0$. We define now $\nu_1 := \kappa'\mu'\nu_0$. By virtue of $y_0 \in \Phi(x_1)$, relation (4.11) gives us $d\big(w_0 - v_0 + y_0, \Psi(x_1)\big) < \mu'\nu_0$. As a result, there is an element $w_1 \in \Psi(x_1)$ for which the inequality $\|w_0 - v_0 + y_0 - w_1\| < \mu'\nu_0$ is fulfilled.

Continuing the current process, let's assume that the points $x_0, x_1, \dots, x_k$ are known. Furthermore, as suggested from the aforementioned arguments, we should include some other points $v_0 \in \Phi(x_0)$, $\dots$, $v_{k-1} \in \Phi(x_{k-1})$ and $w_0 \in \Psi(x_0)$, $\dots$, $w_k \in \Psi(x_k)$ being such that:

- $y_i := z + v_i - w_i \in \Phi(x_{i+1})$;

- $\sup_{\|x-x_i\|\leqslant\nu_i,\zeta\in\Phi(x)}\inf_{\zeta'\in\Psi(x)}\|w_i - v_i + \zeta - \zeta'\| < \mu'\nu_i$, $\nu_i := (\kappa'\mu')^i\nu_0$;
- $\|w_i - v_i + y_i - w_{i+1}\| < \mu'\nu_i = \mu'(\kappa'\mu')^i$;
- $\|x_i - x_{i+1}\| < (\kappa'\mu')^i\kappa' C'$.

Let us now set $\nu_k := (\kappa'\mu')^k\nu_0$. Because of $w_k \in \Phi(x_k)$, we get

$$\inf_{v\in\Phi(x_k)}\sup_{\|x-x_k\|\leqslant\nu_k,\zeta\in\Phi(x)}\inf_{\zeta'\in\Psi(x)}\|w_k - v + \zeta - \zeta'\| \leqslant \sigma_{\Phi,\Psi}(x_k,\nu_k) \leqslant \mu\nu_k$$

as a consequence of (4.9). Hence, we are able to select an element v_k in $\Phi(x_k)$ such that

$$\sup_{\|x-x_k\|\leqslant\nu_k,w\in\Phi(x)}\inf_{w'\in\Psi(x)}\|w_k - v_k + w - w'\| < \mu'\nu_k. \tag{4.12}$$

Denoting $y_k = z + v_k - w_k$, we claim $(x_k, y_k) \in \mathcal{V}$. Indeed, thanks to the triangle inequality

$$\|x_k - \bar{x}\| \leqslant \sum_{i=0}^{k-1}\|x_{k-i} - x_{k-i-1}\| \leqslant \sum_{i=0}^{k-1}(\kappa'\mu')^i\kappa' C' < \frac{\kappa'}{1-\kappa'\mu'}s' < r,$$

which verifies $x_k \in \mathbb{B}(\bar{x}, r)$. On the other hand, repeating the arguments as in the case $k = 0$, we obtain

$$\begin{aligned}\|y_k - \bar{y}\| \leqslant \|z - \bar{z}\| &+ \sup_{\xi\in\Phi(x_k)}\inf_{\xi'\in\Psi(x_k)}\|w_k - v_k + \xi - \xi'\| \\ &+ \inf_{\xi\in\Phi(x_k)}\sup_{\xi'\in\Psi(x_k)}\|(\xi - \bar{y}) - (\xi' - \bar{z})\|.\end{aligned}$$

Consequently, the two relations (4.8) and (4.12) yield

$$\|y_k - \bar{y}\| < s' + \mu'\nu_k + \delta = s' + \mu'(\kappa'\mu')^k\kappa' C' + \delta < \left[1 + (\kappa'\mu')^{k+1}\right]s' + \delta < s.$$

In other words, inclusion $(x_k, y_k) \in \mathcal{V}$ is now clear. Invoking again the supposition concerning regularity property of Φ, we deduce

$$d\big(x_k, \Phi^{-1}(y_k)\big) \leqslant \kappa d\big(y_k, \Phi(x_k)\big) \leqslant \kappa\|y_k - v_k\| = \kappa\|z - w_k\|.$$

Recalling $y_{k-1} = z + v_{k-1} - w_{k-1}$, it holds that

$$d\big(x_k, \Phi^{-1}(y_k)\big) \leqslant \kappa\|w_{k-1} - v_{k-1} + y_{k-1} - w_k\| < \kappa'\mu'\nu_{k-1} = (\kappa'\mu')^k\kappa' C'.$$

Hence, we can update x_{k+1} as a point which belongs to $\Phi^{-1}(y_k)$ and satisfies

$\|x_k - x_{k+1}\| < (\kappa'\mu')^k\kappa'C'$. Involving the evaluation in (4.12), and observing that $\nu_k = (\kappa'\mu')^k\kappa'C'$, $\Psi(x_{k+1})$ must contain an element w_{k+1} fulfilling $\|w_k - v_k + y_k - w_{k+1}\| < \mu'\nu_k$. The construction is completed by induction.

Since $\kappa'\mu' < 1$, an analogous argument as in proof of Proposition 4.2 shows that the sequence (x_k) converges to some $x^* \in X$. Observing $\|z - w_k\| = \|w_{k-1} - v_{k-1} + y_{k-1} - w_k\| < (\kappa'\mu')^{k-1}\nu_0$, we arrive $z \in \Psi(x^*)$ after passing to the limit in the inclusion $w_k \in \Psi(x_k)$. Thus, $d\big(\bar{x}, \Psi^{-1}(z)\big) \leqslant \|\bar{x} - x^*\|$. Nevertheless, according to the construction, triangle inequality gives us

$$\|\bar{x} - x^*\| = \|x_0 - x^*\| \leqslant \sum_{k\geqslant 0} \|x_k - x_{k+1}\| \leqslant \sum_{k\geqslant 0} (\kappa'\mu')^k\kappa'C' = \frac{\kappa'}{1-\kappa'\mu'}C'.$$

This implies $d\big(\bar{x}, \Psi^{-1}(z)\big) \leqslant \frac{\kappa'}{1-\kappa'\mu'}C'$. Because the right-hand side $\frac{\kappa'}{1-\kappa'\mu'}C'$ can be arbitrarily close to $\frac{\kappa}{1-\kappa\mu}C$, we reach to the conclusion (4.10). The proof is done. □

4.2 Convergence of Newton-Type Algorithm with Differentiable Set-Valued Maps

We explore the problem of solving generalized equation in Banach spaces of the form

$$0 \in f(x) + F(x), \tag{4.13}$$

for a C^1 map $f : X \longrightarrow Y$ and closed mapping $F : X \rightrightarrows Y$. Chapter 3 has discussed the applicability of Josephy-Newton method in order to approximate a solution of (4.13). In this section, we shall proceed to another strategy based on set-valued differentiation. Specifically, we focus our consideration on the following scheme

$$0 \in f(x_k) + Df(x_k)(x_{k+1} - x_k) + F(x_k) + H(x_k)(x_{k+1} - x_k) \tag{4.14}$$

corresponding to a map $H : X \longrightarrow \mathcal{PH}(X, Y)$. A typical and important class of such maps is induced by one predeclaring multifunction $\mathcal{H} : X \rightrightarrows \mathcal{L}(X, Y)$ from X into the space $\mathcal{L}(X, Y)$ of linearly continuous maps between X and Y (see, e.g. [5])

$$H(x)(u) = \Big\{Au : A \in \mathcal{H}(x) \subset \mathcal{L}(X, Y)\Big\}.$$

Returning back to the current section, the notion of set-valued differentiability in Definition 2.5 of Section 2.2 is a key point for our analysis. Observe that approximating mappings $H(x_k)$ vary at each step of the iterative process. This motivates us to the following definition.

Definition 4.4. Let $H : X \longrightarrow \mathcal{PH}(X,Y)$ be a given map, and $\Omega \subset X$ be a nonempty open set. Consider some some set-valued map $\Phi : X \rightrightarrows Y$.

(4.4-1). Φ is said to be pointwise strictly differentiable with respective to H on Ω if and only if for any $x \in \Omega$ and $\varepsilon > 0$, there exists $\delta = \delta(x,\varepsilon) > 0$ such that

$$\Phi(z') \subset \Phi(z) + H(x)(z'-z) + \varepsilon \|z'-z\| \mathbb{B};\ \forall z, z' \in \mathbb{B}(x,\delta) \cap \Omega. \tag{4.15}$$

(4.4-2). Φ is differentiable with respective to H uniformly on Ω provided for all $\varepsilon > 0$ there is $\delta = \delta(\varepsilon, \Omega) > 0$ satisfying

$$\Phi(z') \subset \Phi(z) + H(z)(z'-z) + \varepsilon \|z'-z\| \mathbb{B};\ \text{if } z, z' \in \Omega,\ \|z'-z\| \leqslant \delta. \tag{4.16}$$

If $x \in \Omega$ is fixed, then (4.15) covers the concept of $H(x)$-strict differentiability for Φ while (4.16) implies that Φ is $H(x)$-outer differentiable. However, the requirement of uniformity in (4.16) asserts something being stronger.

Associated with a map $H : X \longrightarrow \mathcal{PH}(X,Y)$, we define a new real-valued function by the following formula

$$\Lambda_H(x,x',t) := \sup_{\|u\|\leqslant t} e\big(H(x)(u) + H(x')(-u), 0\big);\ x, x' \in X,\ t \geqslant 0. \tag{4.17}$$

Remark 4.5. It will be shown in Theorem 4.6 that the function Λ_H plays a significant role in the semi-local convergence for the framework of (4.14). Suppose now that both $|H(x)|^+$ and $|H(x')|^+$ are finite. By definition of outer norm we reach to the inclusions $H(x)(u) \subset |H(x)|^+ \|u\| \mathbb{B}_Y$ along with $H(x')(-u) \subset |H(x')|^+ \|u\| \mathbb{B}_Y$. Consequently,

$$e\big(H(x)(u) + H(x')(-u), 0\big) \leqslant \Big(|H(x)|^+ + |H(x')|^+\Big) \|u\|.$$

That is, when $|H(\cdot)|^+$ is bounded, all real-valued functions $\Lambda_H(x,x',\cdot)$ can be majorized by one linear map.

The main results of this section are presented in the two next Theorems 4.6 and 4.10. We begin first with the semi-local version for the convergence of algorithm (4.14).

Theorem 4.6 (semi-local analysis)**.** *Let Ω be an open convex subset of X on which Df is Lipschitz continuous with a modulus $L > 0$. Let $H : X \longrightarrow \mathcal{PH}(X,Y)$ be a map that F is differentiable with respect to H uniformly on Ω. Fix some $x \in \Omega$ with $|H(x)|^+ < +\infty$ and consider a ball $\Omega_x = \bar{\mathbb{B}}(x,r)$ contained into Ω. Suppose that ρ_0 and ε_0 are some positive numbers satisfying*

$$F(z') \subset F(z) + H(z)(z'-z) + \rho_0\|z'-z\|\mathbb{B}_X;\ z', z \in \Omega_x, \|z'-z\| \leqslant \varepsilon_0. \quad (4.18)$$

In addition, we also involve the following assumptions:

(i) it holds that $\tau \in \operatorname{Regmod}\big(\Psi_x, V_{r,s}(\Psi_x)\big)$, where $\Psi_x(\cdot) := Df(x)(\cdot) + H(x)(\cdot)$ and $V_{r,s}(\Psi_x) = \big\{(v,w) : \|v\| \leqslant r, d\big(w, \Psi_x(v)\big) \leqslant s\big\}$;

(ii) $d\big(0, f(x) + F(x)\big) < \min\{s, \tau^{-1}\varepsilon_0\}$;

(iii) it has some (strictly) increasing function $\varrho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that

$$\sigma_{H(z),H(z')}(u,\delta) \leqslant \varrho(\|z-z'\|)\delta, \quad (4.19)$$

for $z, z' \in \bar{\mathbb{B}}(x,r)$, $\|u\| \leqslant r$ and $\delta \leqslant \tau s$;

(iv) we have $\Lambda_H(z,z',t) \leqslant \varphi(t)$ for all $z, z' \in \Omega_x$ and $t \leqslant \varepsilon_0$, where $\varphi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a nondecreasing convex function with $\varphi(0^+) = \limsup_{t\searrow 0}\varphi(t) = 0$;

(v) $\tau[Lr + \varrho(r)] < 1$ and $K_0 = \frac{1}{2}L\varepsilon_0 + \rho_0 + \frac{\varphi(\varepsilon_0)}{\varepsilon_0} < \frac{1}{\tau} - [Lr + \varrho(r)]$.

If either $\frac{1-\tau[Lr+\varrho(r)]}{1-\tau K_0-\tau[Lr+\varrho(r)]}\varepsilon_0 \leqslant r$ or $\frac{1-\tau[Lr+\varrho(r)]}{1-\tau K_0-\tau[Lr+\varrho(r)]}\tau s \leqslant r$ is valid, then there exists a solution x^* of problem (4.13) *such that*

$$\|x - x^*\| \leqslant \frac{1-\tau[Lr+\varrho(r)]}{1-\tau K_0-\tau[Lr+\varrho(r)]}\min\big\{\varepsilon_0, \tau s\big\} \leqslant r. \quad (4.20)$$

Additionally, algorithm (4.14) *generates a sequence (x_k) which starts at $x_0 = x$ and converges R-linearly to x^*, i.e., $\limsup_{k\to\infty}\|x_k - x^*\|^{1/k} < 1$.*

Before proving this theorem, we note that under assumption (iii) the function $|H(\cdot)|^+$ is bounded on $\mathbb{B}(x,r)$. The next technical lemmas will be useful for our proof.

Lemma 4.7. *Keep in mind the hypotheses of Theorem 4.6. For each z and z' in $\mathbb{B}(x,r)$, we set $\Psi_z(\cdot) := Df(z)(\cdot) + H(z)(\cdot)$ and $\Psi_{z'}(\cdot) := Df(z')(\cdot) + H(z')(\cdot)$. Then one has*

$$\sigma_{\Psi_z,\Psi_{z'}}(u,\delta) \leqslant [L\|z-z'\| + \varrho(\|z-z'\|)]\,\delta, \quad (4.21)$$

whenever $u \in r\bar{\mathbb{B}}$ *and* $\delta \leqslant \tau s$.

Proof. Pick $z, z' \in \mathbb{B}(x, r)$ and $u \in r\bar{\mathbb{B}}$. Let u' with $\|u' - u\| \leqslant \delta \leqslant \tau s$ and let $w \in \Psi_z(u)$, $w' \in \Psi_{z'}(u)$, $v \in \Psi_z(u')$, $v' \in \Psi_{z'}(u')$. Based on triangle inequality, it holds that

$$\begin{aligned} \|w' - w + v - v'\| &\leqslant \|Df(z')(u) - Df(z)(u) + Df(z)(u') - Df(z')(u')\| \\ &\quad + \|\omega' - \omega + \zeta - \zeta'\| \\ &\leqslant L\|z - z'\|\|u - u'\| + \|\omega' - \omega + \zeta - \zeta'\| \\ &\leqslant L\|z - z'\|\delta + \|\omega' - \omega + \zeta - \zeta'\|, \end{aligned}$$

in which $\omega' = w' - Df(z')(u) \in H(x')(u)$, $\omega = w - Df(z)(u) \in H(x)(u)$, $\zeta = v - Df(z)(u') \in H(x)(u')$ and $\zeta' = v' - Df(z')(u') \in H(x')(u')$. Thus, according to Definition 4.1, we obtain

$$\sigma_{\Psi_z,\Psi_{z'}}(u, \delta) \leqslant L\|z - z'\|\delta + \sigma_{H(x),H(x')}(u, \delta) \leqslant L\|z - z'\|\delta + \varrho(\|z - z'\|)\delta.$$

This completes the proof of Lemma 4.7. □

Lemma 4.8. *Let* L, τ, ρ_0, λ *and* r *be in the statement of Theorem 4.6. Define some parameters* $\gamma_1 = \frac{1}{2}\tau L$, $\gamma_2 = \rho_0\tau$ *and* $\gamma_3 = \tau[Lr + \varrho(r)] < 1$. *Let* $h(t) = \frac{1}{1-\gamma_3}\big(\gamma_1 t^2 + \gamma_2 t + \tau\varphi(t)\big)$, *where* φ *is the function appeared in Theorem 4.6. Then, under the initial condition* $h(\alpha_0) \leqslant \alpha_0$, *the recurrence* $\alpha_{k+1} = h(\alpha_k)$ *generates a sequence converging linearly.*

Proof. If $h(\alpha_0) = \alpha_0$, then it is easy to see that $\alpha_k = \alpha_0$ for all k. In this case, the conclusion is straightforward. Otherwise, suppose $\alpha_0 = qh(\alpha_0)$ with $q \in (0, 1)$. φ is a convex function, so is h. Thus, the function $h_1(t) = \frac{h(t)}{t}$ is increasing (see e.g. [42]), which implies $h(t) \leqslant qt$ for $t \in [0, \alpha_0]$. Note that $\alpha_1 = h(\alpha_0) < \alpha_0$, thus $\alpha_2 = h(\alpha_1)$ makes sense, and $\alpha_2 \leqslant q\alpha_1$. Repeating this procedure, the sequence (α_k) is well-defined and obeys the relation $\alpha_{k+1} \leqslant q\alpha_k$. The proof is done. □

Proof of Theorem 4.6. We separate the proof in several parts.

Step1: Approximation sequence.

At the beginning, let's set $x_0 = x$, $\Psi_0 = \Psi_x$, $r_0 = r$, $s_0 = s$ and $\mathcal{V}_0 = V_{r,s}(\Psi_x)$. Then the mapping Ψ_0 is metrically regular on the set $\mathcal{V}_0$ together with a modulus $\tau_0 = \tau$. Assumption (ii) permits us to select in $F(x_0)$ an element y_0 such that $\|-f(x_0) - y_0\| < \min\{s, \tau^{-1}\varepsilon_0\}$. Denoting $z_0 = -f(x_0) - y_0$, we have

$$d\big(z_0, \Psi_0(0)\big) = d\big(z_0, H(x)(0)\big) = \|z_0\| < s.$$

This means $(0, z_0) \in \mathcal{V}_0$, which implies

$$d\big(0, \Psi_0^{-1}(z_0)\big) \leqslant \tau_0 d\big(z_0, \Psi_0(0)\big) = \tau_0 \|z_0\| < \min\big\{\tau s, \varepsilon_0\big\}.$$

Therefore, the set $\Psi_0^{-1}(z_0)$ contains an element u_0 such that $\|u_0\| < \min\big\{\tau s, \varepsilon_0\big\}$. Define $x_1 = x_0 + u_0$, we get

$$-f(x_0) - y_0 = z_0 \in \Psi_0(x_1 - x_0) = Df(x_0)(x_1 - x_0) + H(x_0)(x_1 - x_0).$$

Because of $y_0 \in F(x_0)$, we reach to the following inclusion

$$0 \in f(x_0) + Df(x_0)(x_1 - x_0) + H(x_0)(x_1 - x_0) + F(x_0).$$

In other words, x_1 is produced by scheme (4.14) and

$$\|x_1 - x_0\| = \|u_0\| < \min\big\{\tau s, \varepsilon_0\big\}.$$

Put $\alpha_0 = \min\big\{\tau s, \varepsilon_0\big\}$. We include the function h as mentioned in Lemma 4.8. Then, the hypothesis (v) gives us $\frac{h(\varepsilon_0)}{\varepsilon_0} = \frac{1}{1-\tau[Lr+\varrho(r)]}\tau K_0 < 1$. Furthermore, the function $t \longmapsto t^{-1}h(t)$ is increasing on the interval $(0, \alpha_0]$ (Lemma 4.8). Thus,

$$h(\alpha_0) \leqslant \varepsilon_0^{-1} h(\varepsilon_0)\alpha_0 = q\alpha_0, \quad q = \frac{1}{1 - \tau[Lr + \varrho(r)]}\tau K_0.$$

Invoking Lemma 4.8, the sequence (α_k) given by $\alpha_{i+1} = h(\alpha_i)$ is well-defined and $\alpha_{k+1} \leqslant q\alpha_k$. In particular, we have $\alpha_k \leqslant q^k \alpha_0$ for $k = 0, 1, \ldots$

Let's proceed to the induction process. Suppose that the iterations $x_0, x_1, \ldots, x_k$ are given. As suggested in the preceding part, we require the following relations are fulfilled:

- $0 \in f(x_i) + Df(x_i)(x_{i+1} - x_i) + H(x_i)(x_{i+1} - x_i) + F(x_i)$, $i \leqslant k - 1$;
- $\|x_{i+1} - x_i\| < \alpha_i \leqslant q^i \alpha_0$, $i = 0, \ldots, k - 1$.

Observe that all terms x_i are in the ball $\mathbb{B}(x, r)$. Indeed, this is a consequence of subsequent estimations

$$\sum_{i=0}^{k-1} \alpha_i \leqslant \sum_{i=0}^{k-1} q^i \alpha_0 < \frac{1}{1-q}\alpha_0 = \frac{1 - \tau[Lr + \varrho(r)]}{1 - \tau K_0 - \tau[Lr + \varrho(r)]} \min\big\{\varepsilon_0, \tau s\big\} \leqslant r.$$

If $x_{k-1} = x_k$, we simply set $x_{k+1} = x_k$. Otherwise, we write $u_i = x_{i+1} - x_i$ for $i \leqslant k - 1$ and denote $\Psi_k(\cdot) = Df(x_k)(\cdot) + H(x_k)(\cdot)$. Thanks to Lemma 4.7, one

has
$$\sigma_{\Psi_x,\Psi_k}(u,\delta) \leqslant [L\|x-x_k\| + \varrho(\|x-x_k\|)]\,\delta = \beta_k\delta,$$
where the parameter $\beta_k = L\|x-x_k\| + \varrho(\|x-x_k\|)$ satisfies $\beta_k \leqslant Lr + \varrho(r) < \tau^{-1}$. Based on Proposition 4.2, if $(u,z) \in X \times Y$ fulfills

$$\|u\| + \frac{\tau}{1-\beta_k\tau} d\big(z, \Psi_k(u)\big) < r, d\big(z, \Psi_k(u)\big) < s, \tag{4.22}$$

then the estimation below

$$d\big(u, \Psi_k^{-1}(z)\big) \leqslant \tau_k d\big(z, \Psi_k(u)\big), \tau_k = \frac{\tau}{1-\beta_k\tau} \tag{4.23}$$

follows immediately.
We shall generate x_{k+1} through the scheme (4.14). By inductive hypothesis, there exists a point $y_{k-1} \in F(x_{k-1})$ for which $w_{k-1} \in H(x_{k-1})(u_{k-1})$ and

$$w_{k-1} = -f(x_{k-1}) - Df(x_{k-1})(u_{k-1}) - y_{k-1}.$$

Since $\alpha_{k-1} \leqslant q^{k-1}\alpha_0 \leqslant \varepsilon_0$, (4.18) can be applied to $z' = x_{k-1}$ and $z = x_k$. Consequently, there are some elements $y_k \in F(x_k)$ and $w'_{k-1} \in H(x_k)(-u_{k-1})$ such that
$$y_{k-1} - y_k - w'_{k-1} \in \rho_0\|u_{k-1}\|\,\mathbb{B}.$$
Define $z_k = -f(x_k) - y_k$ and $u_k^* = 0$, we claim that $(u,z) = (u_k^*, z_k)$ obeys (4.22). In fact, it is possible to represent z_k into another form as follows

$$z_k = [-f(x_k) + f(x_{k-1}) + Df(x_{k-1})(u_{k-1})] + [y_{k-1} - y_k - w'_{k-1}] + w_{k-1} + w'_{k-1}.$$

Remind $y_{k-1} - y_k - w'_{k-1} \in \rho_0\|u_{k-1}\|\,\mathbb{B}$, one has $\left\|y_{k-1} - y_k - w'_{k-1}\right\| < \rho_0\alpha_{k-1}$. In addition, because of $w_{k-1} + w'_{k-1} \in H(x_{k-1})(u_{k-1}) + H(x_k)(-u_{k-1})$, assumption (iv) yields
$$\left\|w_{k-1} + w'_{k-1}\right\| \leqslant \Lambda_H(x_{k-1}, x_k, \alpha_{k-1}) \leqslant \varphi(\alpha_{k-1}).$$
Because Df is Lipschitz continuous, the Taylor's expansion for f at x_{k-1} gives us

$$\begin{aligned}
&\|-f(x_k) + f(x_{k-1}) + Df(x_{k-1})(u_{k-1})\| \\
&\quad = \left\|\int_0^1 [Df(x_{k-1} + tu_{k-1})(u_{k-1}) - Df(x_{k-1})](u_{k-1})\,dt\right\| \\
&\quad \leqslant \int_0^1 L\|u_{k-1}\|^2 t\,dt < \frac{1}{2}L\alpha_{k-1}^2.
\end{aligned}$$

Thus, we reach to the following estimation for $d\big(z_k, \Psi_k(0)\big)$

$$d\big(z_k, \Psi_k(0)\big) = d\big(z_k, H(x_k)(0)\big) = \|z_k\| < \frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1}).$$

Let us verify that

$$\frac{\tau}{1-\beta_k\tau}\left(\frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1})\right) < r, \tag{4.24a}$$

$$\frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1}) < s. \tag{4.24b}$$

Since $\beta_k \leqslant Lr + \varrho(r)$, the left-hand side of (4.24a) is dominated by

$$\frac{\tau}{1-\gamma_3}\left(\frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1})\right) = h(\alpha_{k-1}) = \alpha_k \leqslant \alpha_0.$$

Note that $\alpha_0 = \min\big\{\varepsilon_0, \tau s\big\} < r$, we derive (4.24a). Proof of (4.24b) is similar to the one of (4.24a), since $\alpha_0 \leqslant \tau s$ and

$$\frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1}) = \frac{1-\gamma_3}{\tau}h(\alpha_{k-1}) = \frac{1-\gamma_3}{\tau}\alpha_k \leqslant \frac{1-\gamma_3}{\tau}\alpha_0.$$

According to (4.24a) and (4.24b), we can show that $u = u_k^* = 0$ and $z = z_k$ fulfill (4.22). Hence, (4.23) implies

$$d\big(0, \Psi_k^{-1}(z_k)\big) \leqslant \tau_k d\big(z_k, \Psi_k(0)\big) < \frac{\tau}{1-\beta_k\tau}\left(\frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1})\right).$$

As a result, there is an element $u_k \in \Psi_k^{-1}(z_k)$ such that

$$\|u_k\| < \frac{\tau}{1-\beta_k\tau}\left(\frac{1}{2}L\alpha_{k-1}^2 + \rho_0\alpha_{k-1} + \varphi(\alpha_{k-1})\right) \leqslant h(\alpha_{k-1}) = \alpha_k.$$

Let us set $x_{k+1} = x_k + u_k$. As similar to the case $k = 0$, we conclude that x_{k+1} is obtained through algorithm (4.14). By inductive principle, the sequence (x_k) is well-defined.

Step2: Convergence.

Recalling $q \in (0,1)$, the series $\sum_{k\geqslant 0}\alpha_k$ converges. Because $\|x_k - x_{k+1}\|$ is majorized by α_k, (x_k) is thereby a Cauchy sequence. Let $x^* = \lim\limits_{k\to\infty} x_k$, we find

$$\|x_k - x^*\| \leqslant \sum_{i\geqslant k}\|x_i - x_{i+1}\| \leqslant \sum_{i\geqslant k}\alpha_i \leqslant \sum_{i\geqslant k} q^i\alpha_0 = \frac{q^k}{1-q}\alpha_0. \tag{4.25}$$

R-linear convergence and (4.20) follows immediately from (4.25). To show that x^* solves the initial problem (4.13), we pass to the limit in (4.14) and include the inclusion $H(x_k)(u) \subset |H(x_k)|^+ \|u\|$ there. □

Remark 4.9. In the manuscript [7], the authors have proved a statement which subsumes as a particular case to Theorem 4.6. The corresponding one in [7] is based on a stronger assumption than the aforementioned theorem, which confine ϱ to a linear map (i.e., $\varrho(t) = \lambda t$ for some $\lambda > 0$). However, the linearity property of $\varrho(\cdot)$ does not play any extraordinary role. Indeed, the importance of function $\varrho(\cdot)$ is that, it permits us to control the magnitude of measuring quantity $\sigma_{H(x),H(x')}(u,\delta)$ when x' varies around x. As we have seen above, for such a purpose, it is sufficient to exploit only the monotonicity property imposed on $\varrho(\cdot)$.

The rest of the current section is devoted to study the local behavior of algorithm (4.14). Precisely, we have the theorem below.

Theorem 4.10 (local convergence)**.** *Suppose that problem* (4.13) *admits* $x^* \in X$ *as a solution. Let* $H : X \longrightarrow \mathcal{PH}(X,Y)$ *be a given map so that* F *is pointwise strictly differentiable with respect to* H *at* x^**. Additionally, we assume that there are two increasing continuous functions* $\rho, \varrho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ *and a positive number* $\bar{r} > 0$ *for which the following assumptions hold:*

(A1) $F(x') \subset F(x) + H(x)(x'-x) + \rho(\|x'-x\|)\mathbb{B}$ *whenever* $x, x' \in \mathbb{B}(x^*, \bar{r})$*,*

(A2) $\sigma_{H(x^*),H(x)}(u,\varepsilon) \leqslant \varrho\left(\|x-x^*\|\right)\varepsilon$*, for all* $u \in \bar{r}\bar{\mathbb{B}}$ *and* $\varepsilon \leqslant \bar{r}$*,*

(A3) $\lambda^* = \limsup_{t \to 0} \left(t^{-1}\rho(t)\right) < +\infty$.

On the other hand, suppose that:

- $\tau^* = \operatorname{Reg} \Phi_{x^*}(0,0) < +\infty$ *for* $\Phi_{x^*}(\cdot) := Df(x^*)(\cdot) + H(x^*)(\cdot)$*;*
- $\tau^*\left(\lambda^* + \varrho(0)\right) < 1$*;*
- Df *is Lipschitz continuous while* $|H(\cdot)|^+$ *is finite on the ball* $\mathbb{B}(x^*, \bar{r})$*.*

Then, there exists a constant $0 < \alpha < \bar{r}$ *having the property below. For any guess point* $x \in \mathbb{B}(x^*, \alpha)$*, the framework of* (4.14) *generates a sequence* (x_k) *such that* $x_0 = x$ *and* $x_k \to x^*$ *at least linearly.*

We will need some preparatory lemmas to prove this theorem.

Lemma 4.11. *Keeping all statements in the hypotheses of Theorem 4.10. For a given point* $\bar{x} \in \mathbb{B}(x^*, \alpha)$*, we set* $\Psi_{x^*}(\cdot) = Df(x^*)(\cdot - \bar{x}) + H(x^*)(\cdot - \bar{x})$*,* $\Psi_{\bar{x}}(\cdot) = Df(\bar{x})(\cdot - \bar{x}) + H(\bar{x})(\cdot - \bar{x})$*. Suppose* Df *is Lipschitz continuous on* $\mathbb{B}(x^*, \bar{r})$ *with a modulus* L*, then we have the estimation*

$$\sigma_{\Psi_{x^*},\Psi_{\bar{x}}}(x, \varepsilon) \leqslant [L \|\bar{x} - x^*\| + \varrho(\|\bar{x} - x^*\|)]\, \varepsilon, \forall x \in \mathbb{B}(\bar{x}, \bar{r}), \varepsilon \leqslant \bar{r}. \tag{4.26}$$

Proof. Using triangle inequality in Y, the definition of $\sigma_{\Psi_{x^*},\Psi_{\bar{x}}}(u, \varepsilon)$ provides us

$$\begin{aligned}
\sigma_{\Psi_{x^*},\Psi_{\bar{x}}}(x, \varepsilon) \leqslant & \sup_{\|x'-x\| \leqslant \varepsilon} \|[Df(\bar{x}) - Df(x^*)]\,(x - \bar{x}) + [Df(x^*) - Df(\bar{x})]\,(x' - \bar{x})\| \\
& + \sup_{\zeta \in H(\bar{x})(x-\bar{x})} \inf_{\zeta' \in H(x^*)(x-\bar{x})} \sup_{\substack{\|x'-x\| \leqslant \varepsilon, \\ \xi \in H(x^*)(x'-\bar{x})}} \inf_{\xi' \in H(\bar{x})(x'-\bar{x})} \|\zeta - \zeta' + \xi - \xi'\| \\
& \leqslant \sup_{\|x'-x\| \leqslant \varepsilon} \|[Df(\bar{x}) - Df(x^*)]\,(x - x')\| + \sigma_{H(x^*),H(\bar{x})}(x - \bar{x}, \varepsilon) \\
& \leqslant \sup_{\|x'-x\| \leqslant \varepsilon} \|Df(\bar{x}) - Df(x^*)\|\,\|x - x'\| + \sigma_{H(x^*),H(\bar{x})}(x - \bar{x}, \varepsilon) \\
& \leqslant L \|\bar{x} - x^*\|\, \varepsilon + \varrho(\|\bar{x} - x^*\|)\, \varepsilon.
\end{aligned}$$

Hence, the proof is done. □

Lemma 4.12. *Suppose that* Φ_{x^*} *is metrically regular on some neighborhood* $r\mathbb{B} \times s\mathbb{B}$ *with modulus* τ*. Let* $0 < \delta < r$ *and* $0 < \delta' < s$ *be given. For* $\bar{x} \in \bar{\mathbb{B}}(x^*, r - \delta)$*, we consider the mapping* $\Phi_{\bar{x}}(\cdot) = Df(x^*)(\cdot - \bar{x}) + H(x^*)(\cdot - \bar{x})$*. If* $\bar{z} \in \Phi_{\bar{x}}(x^*)$ *satisfies* $\|\bar{z}\| \leqslant s - \delta'$*, then one has* $\tau \in \operatorname{Regmod}\big(\Phi_{\bar{x}}, V_{\delta,\delta'}\big)$*, where* $V_{\delta,\delta'} = \mathbb{B}(x^*, \delta) \times \mathbb{B}(\bar{z}, \delta')$*.*

Proof. Pick $(x, z) \in V_{\delta,\delta'}$. By the definition of $\Phi_{\bar{x}}$, it is easy to check that

$$\Phi_{\bar{x}}(x) = \Phi_{x^*}(x - \bar{x}), \Phi_{\bar{x}}^{-1}(z) = \Big\{\bar{x} + u : z \in \Phi_{x^*}(u)\Big\} = \bar{u} + \Phi_{x^*}^{-1}(z).$$

Since $\|x - \bar{x}\| \leqslant \|x - x^*\| + \|x^* - \bar{x}\| < r$ and $\|z\| \leqslant \|x - \bar{z}\| + \|\bar{z}\| < s$, the metric regularity property for Φ_{x^*} can be applied to $(x - \bar{x}, z)$. Thus, we get

$$d\big(x - \bar{x}, \Phi_{x^*}^{-1}(z)\big) \leqslant \tau d\big(z, \Phi_{x^*}(x - \bar{x})\big) = \tau d\big(z, \Phi_{\bar{x}}(x)\big).$$

Equivalently, the latter is rewritten as follows $d\big(x, \Phi_{\bar{x}}^{-1}(z)\big) \leqslant \tau d\big(z, \Phi_{\bar{x}}(x)\big)$. Hence, we reach to the conclusion of Lemma 4.12. □

Lemma 4.13. *Keep in mind all assumptions of Theorem 4.10. Let* $x \in \mathbb{B}(x^*, \bar{r})$

and $\|u\| < \bar{r}$*. If* r *is a number such that* $\|x - x^*\| < r < \bar{r}$*, then one has*

$$H(x)(u) \subset \left(|H(x^*)|^+ + \varrho(r)\right) \|u\| \, \bar{\mathbb{B}}. \tag{4.27}$$

As a consequence, the following estimation holds

$$\inf_{\upsilon \in H(x^*)(u)} \sup_{\omega \in H(x)(u)} \|(\omega - \bar{\omega}) - (\upsilon - \bar{\upsilon})\| \leqslant \left(2\,|H(x^*)|^+ + \varrho(r)\right) \|u\| + \|\bar{\omega} - \bar{\upsilon}\| \,. \tag{4.28}$$

Proof. Fix some constant b with $\|u\| < b < \bar{r}$, and take $\varepsilon = (\|u\| + b)/2$. Since $\varepsilon < \bar{r}$, we have $\sigma_{H(x^*),H(x)}(u, \varepsilon) \leqslant \varrho(\|x - x^*\|)\varepsilon < \varrho(r)\varepsilon$. If $w \in H(x)(u)$, then

$$\inf_{\upsilon \in H(x^*)(u)} \sup_{\substack{\|u'-u\| \leqslant \varepsilon, \\ \zeta \in H(x^*)(u')}} \inf_{\zeta' \in H(x)(u')} \|w - \upsilon + \zeta - \zeta'\| \leqslant \sigma_{H(x^*),H(x)}(u, \varepsilon) < \varrho(r)\varepsilon.$$

Thus, there is $w' \in H(x^*)(u)$ such that

$$\sup_{\|u'-u\| \leqslant \varepsilon, \zeta \in H(x^*)(u')} \inf_{\zeta' \in H(x)(u')} \|w - w' + \zeta - \zeta'\| < \varrho(a)\varepsilon.$$

By replacing u' with 0, and using the fact $H(x)(0) = H(x^*) = \{0\}$, we deduce $\|w - w'\| < \varrho(r)\varepsilon$. Furthermore, it follows from the inclusion $w' \in H(x^*)(u)$ that $\|w'\| \leqslant |H(x^*)|^+ \|u\|$. The latter implies $\|w\| \leqslant |H(x^*)|^+ \|u\| + \varrho(r)\varepsilon$. But b can be arbitrarily close to $\|u\|$ without depending on w, we conclude $\|w\| \leqslant \left(|H(x^*)|^+ + \varrho(r)\right) \|u\|$. In other words, inclusion (4.27) is proved. At last, (4.28) is inferred from (4.27). □

Proof of Theorem 4.10. Let $L \in \operatorname{Lipmod}\left(Df, \mathbb{B}(x^*, r)\right)$ and let $\nu^* = \|Df(x^*)\| + |H(x^*)|^+$. Choose some parameters $\tau > \tau^*$ and $\lambda > \lambda^*$ such that $\tau[\lambda + \varrho(0)] < 1$. Alternatively, we take $r \in (0, \bar{r})$ and $s > 0$ so that Ψ^* is metrically regular with modulus τ on the neighborhood $V = r\mathbb{B} \times s\mathbb{B}$. Next, one searches a value $0 < \alpha < r/2$ satisfying simultaneously three constraints (4.29a), (4.29b) and (4.29c) below

$$\rho(t) \leqslant \lambda t \quad \text{when } 0 \leqslant t \leqslant \alpha; \tag{4.29a}$$

$$\tau \left(\frac{3}{2} L\alpha + \varrho(\alpha) + \lambda\right) < 1; \tag{4.29b}$$

$$\left[L\alpha + 3\varrho(\alpha) + 4\,|H(x^*)|^+ + \nu^*\right] \alpha + 2\left(\frac{1}{2} L\alpha^2 + \rho(\alpha)\right) < s. \tag{4.29c}$$

Let x be in the ball $\mathbb{B}(x^*, \alpha)$. The case $x = x^*$ is trivial. Otherwise, we set $x_0 = x$

and denote $u_0 = x_0 - x^*$, $\Phi_0(\cdot) = Df(x^*)(\cdot - x_0) + H(x^*)(\cdot - x_0)$. Define some parameters $\alpha_0 = \|u_0\|$ and $\mu_0 = L\alpha_0 + \varrho(\alpha_0)$. By assumptions of Theorem 4.10

$$-f(x^*) \in F(x^*) \subset F(x_0) + H(x_0)(-u_0) + \rho(\|u_0\|)\mathbb{B},$$

which permits us to select $y_0 \in F(x_0)$ and $w_0 \in H(x_0)(-u_0)$ such that

$$-f(x^*) \in y_0 + w_0 + \rho(\alpha_0)\mathbb{B}.$$

Because of $\alpha_0 = \|u_0\| < r$, one has $\sigma_{H(x^*),H(x_0}(-u_0, \alpha_0) \leqslant \varrho(\alpha_0)\alpha_0$. So, we derive from the relation $w_0 \in H(x_0)(-u_0)$ that

$$\inf_{\xi \in H(x^*)(-u_0)} \sup_{\|u+u_0\| \leqslant \alpha_0, \upsilon \in H(x^*)(u)} \inf_{\upsilon' \in H(x_0)(u)} \|w_0 - \xi + \upsilon - \upsilon'\| \leqslant \varrho(\alpha_0)\alpha_0.$$

Observing $\alpha_0 < \alpha$, there is an element $w_0^* \in H(x^*)(u_0)$ with

$$\sup_{\|u+u_0\| \leqslant \alpha_0, \upsilon \in H(x^*)(u)} \inf_{\upsilon' \in H(x_0)(u)} \|w_0 - w_0^* + \upsilon - \upsilon'\| < \varrho(\alpha)\alpha.$$

Hence, after substituting $u = 0$ and using the fact $H(x^*)(0) = H(x_0)(0) = \{0\}$, we get $\|w_0 - w_0^*\| < \varrho(\alpha)\alpha$.

Next, we denote $\Psi_0(\cdot) = Df(x_0)(\cdot - x_0) + H(x_0)(\cdot - x_0)$ and consider some new points $z_0 = -f(x_0) - y_0$, $\bar{z}_0 = -Df(x_0)(u_0) + w_0$ and $z_0^* = -Df(x^*)(u_0) + w_0^*$. By virtue of $w_0 \in H(x_0)(-u_0)$, $\bar{z}_0$ is in $\Psi_0(x^*)$. Remind $w_0^* \in H(x^*)(-u_0)$, the following estimation is fulfilled

$$\|z_0^*\| \leqslant \|Df(x^*)\| \, \|u_0\| + |H(x^*)|^+ \, \|u_0\| = \nu^*\alpha_0.$$

Thanks to (4.29c), it is easy to see that $\nu^*\alpha_0 \leqslant \nu^*\alpha < s$. Using Lemma 4.12, the mapping Φ_0 is metrically regular with modulus τ on the neighborhood $V_0 = \mathbb{B}(x^*, \alpha) \times \mathbb{B}(z_0^*, s - \nu^*\alpha_0)$ of (x^*, z_0^*). We shall apply Proposition 4.3 to produce the next iteration x_1.

For this goal, define with respect to $x \in \mathbb{B}(x^*, \alpha)$ the quantity

$$\gamma(x, x_0) := \inf_{\upsilon \in \Phi_0(x)} \sup_{\upsilon' \in \Psi_0(x)} \|(\upsilon' - \bar{z}_0) - (\upsilon - z_0^*)\|.$$

Then, it is possible to verify the following estimation

$$\gamma(x, x_0) \leqslant \|[Df(x_0) - Df(x^*)](x - x^*)\|$$

$$+ \inf_{z\in H(x^*)(x-x_0)} \sup_{w\in H(x)(x-x_0)} \|(w-w_0)-(z-w_0^*)\| .$$

Nevertheless, for $x \in \mathbb{B}(x^*, \alpha)$, Lemma 4.13 gives us

$$\begin{aligned} &\inf_{z\in H(x^*)(x-x_0)} \sup_{w\in H(x)(x-x_0)} \|(w-w_0)-(z-w_0^*)\| \\ &\qquad \leqslant \big(2\,|H(x^*)|^+ + \varrho(\alpha)\big)\,\|x-x_0\| + \|w_0-w_0^*\| . \end{aligned} \tag{4.30}$$

Observing $\|x-x_0\| \leqslant \|x-x^*\| + \|x^*-x_0\| < 2\alpha$, so the Lipschitz continuity of Df and (4.30) yield

$$\begin{aligned} \gamma(x,x_0) &\leqslant \alpha L\,\|x_0-x^*\| + 2\alpha\left[2\,|H(x^*)|^+ + \varrho\,(\alpha)\right] + \|w_0-w_0^*\| \\ &= \alpha\alpha_0 L + 2\alpha\left[2\,|H(x^*)|^+ + \varrho\,(\alpha)\right]\alpha + \varrho(\alpha)\alpha,\ x\in\mathbb{B}(x^*,\alpha). \end{aligned} \tag{4.31}$$

Let $x \in \mathbb{B}(x^*, \alpha)$ and $\varepsilon \leqslant \alpha$, we obtain from Lemma 4.11 that

$$\sigma_{\Phi_0,\Psi_0}(x,\varepsilon) \leqslant [L\,\|x_0-x^*\| + \varrho\,(\|x_0-x^*\|)]\,\varepsilon = \mu_0\varepsilon.$$

According to (4.29b), $\tau\mu_0$ is evident less than 1. Thus, if the group of estimates below is valid

$$\begin{cases} \frac{\tau}{1-\tau\mu_0}\,\|z-\bar{z}_0\| < \alpha, \\ \sup_{x\in\mathbb{B}(x^*,\alpha)} \gamma(x,x_0) + (1+\tau\mu_0)\,\|z-\bar{z}_0\| < s-\nu^*\alpha_0, \end{cases} \tag{4.32}$$

then Proposition 4.3 implies

$$d\big(x^*, \Psi_0^{-1}(z)\big) \leqslant \tau_0 d\big(z, \Psi_0(x^*)\big) \leqslant \tau_0\,\|z-\bar{z}_0\|\,,\ \tau_0 = \frac{1}{1-\tau\mu_0}\tau. \tag{4.33}$$

We are now going to claim that $z = z_0$ satisfies (4.32), and then apply (4.33). Indeed, thanks to the triangle inequality, $\|z_0-\bar{z}_0\|$ is majorized by

$$\begin{aligned} \|z_0-\bar{z}_0\| &= \|-f(x_0)-y_0+Df(x_0)(u_0)-w_0\| \\ &\leqslant \|f(x^*)-f(x_0)-Df(x_0)(-u_0)\| + \|-f(x^*)-y_0-w_0\| . \end{aligned}$$

Using the Taylor's expansion for f at center x_0, we obtain

$$\begin{aligned} &\|f(x^*)-f(x_0)-Df(x_0)(-u_0)\| \\ &\qquad = \left\| \int_0^1 \left[Df(tx^*+(1-t)x_0)-Df(x_0)\right](-u_0)\,dt \right\| \end{aligned}$$

$$\leqslant \int_0^1 Lt \|x^* - x_0\| \|u_0\| \; dt = \frac{1}{2}L\alpha_0^2.$$

Recalling $-f(x^*) \in y_0 + w_0 + \rho(\alpha_0)\mathbb{B}$, the term $\|-f(x^*) - y_0 - w_0\|$ is less than $\rho(\alpha_0)$. Combining these arguments, we arrive

$$\|z_0 - \bar{z}_0\| < \frac{1}{2}L\alpha_0^2 + \rho(\alpha_0) \leqslant \frac{1}{2}L\alpha_0^2 + \lambda\alpha_0. \tag{4.34}$$

Therefore, three relations (4.29b), (4.29c) and (4.31) ensure that $z = z_0$ satisfies the property described by (4.32).
Letting $z = z_0$ in (4.33), and invoking (4.34), the evaluation

$$d\left(x^*, \Psi_0^{-1}(z_0)\right) < \frac{\tau}{1 - \tau[L\alpha_0 + \varrho(\alpha_0)]}\left(\frac{1}{2}L\alpha_0^2 + \lambda\alpha_0\right)$$

is fulfilled. As a consequence, we are able to select an element $x_1 \in \Psi_0^{-1}(z_0)$ such that

$$\alpha_1 = \|x^* - x_1\| < \frac{\tau}{1 - \tau[L\alpha_0 + \varrho(\alpha_0)]}\left(\frac{1}{2}L\alpha_0^2 + \lambda\alpha_0\right). \tag{4.35}$$

Taking into account (4.29b), the assignment $\psi(t) = \frac{1}{1-\tau[Lt+\varrho(t)]}\left(\frac{1}{2}\tau Lt + \tau\lambda\right)$ defines a function from the interval $[0, \alpha]$ into $[0, 1)$. Using this function, (4.35) is rewritten as follows

$$\|x^* - x_1\| < \psi(\|x^* - x_0\|) \|x^* - x_0\|. \tag{4.36}$$

To see that x_1 is generated by (4.14), we recall that $z_0 \in \Psi_0(x_1)$. Because of $z_0 = -f(x_0) - y_0$ and $\Psi_0 = Df(x_0)(\cdot - x_0) + H(x_0)(\cdot - x_0)$, the inclusion $y_0 \in F(x_0)$ give us

$$0 \in f(x_0) + Df(x_0)(x_1 - x_0) + H(x_0)(x_1 - x_0) + F(x_0).$$

Observe that (4.36) implies $\|x^* - x_1\| < \psi(\alpha) \|x^* - x_0\| < \alpha$. This allows us to take x_1 as the new starting point instead of x_0, and continue the construction. Repeating this process, we obtain the sequence (x_k) satisfying (4.14) and

$$\|x^* - x_{k+1}\| \leqslant \psi\left(\|x^* - x_k\|\right) \|x^* - x_k\| \leqslant \psi\left(\alpha\right) \|x^* - x_k\|. \tag{4.37}$$

Linear convergence for (x_k) follows directly from (4.37). The proof is thereby completed. □

Remark 4.14. As in Theorem 4.6, the function $\varrho(\cdot)$ also controls the growth of measure quantities $\sigma_{H(x^*),H(x)}(\cdot,\cdot)$, which guarantees the stability of the metric regularity property when the current iteration is nearby x^*. Such a property itself

permits to obtain x_{k+1} from x_k as well as to give an estimation for $\|x_{k+1} - x_k\|$. The presupposition (A2) concerning with $\varrho(\cdot)$ might be acceptable in some situations of applications, e.g., when H is determined through some set-valued map from X to $\mathcal{L}(X, Y)$.
On the other side, the rate of convergence for the approximating sequence (x_k) seems to be mostly induced by the behaviour of $\rho(\cdot)$ around 0. If F is differentiable with respect to H uniformly on a neighborhood Ω^* of x^*, then it is sufficient to take $\rho(\cdot)$ as a linear function $\rho(t) = \varepsilon t$ (and choose $\bar{r} \leqslant \delta$ small enough, where $\delta = \delta(\varepsilon, \Omega^*)$ occurred in Definition 4.4). In general, assume that F is pointwise H-strictly differentiable around x^* while the assertion (A2) of Theorem 4.10 is involved. Using (A2), $\sigma_{H(x^*),H(x)}(0, \epsilon) \leqslant \varrho(\|x - x^*\|)\epsilon$, so by a similar technique as in the proof of Lemma 4.13, we can establish that

$$H(x^*)(u) \subset H(x)(u) + 2\varrho(\|x - x^*\|)\epsilon\mathbb{B},$$

with $\|u\| < \epsilon < \bar{r}$ and $\|x - x^*\| < \bar{r}$. Consequently, for x and x' being sufficiently close to x^*, one has

$$\begin{aligned} F(x') &\subset F(x) + H(x^*)(x' - x) + \varepsilon \|x' - x\| \mathbb{B} \\ &\subset F(x) + H(x)(x' - x) + [2\varrho(\|x - x^*\|) + \varepsilon \|x' - x\|]\mathbb{B} \\ &\subset F(x) + H(x)(x' - x) + [2\varrho(\|x - x'\| + \bar{r}) + \varepsilon \|x' - x\|]\mathbb{B}. \end{aligned}$$

Taking $\rho(t) := 2\varrho(t + \bar{r}) + \varepsilon t$, (A1) follows.

According to the proof of Theorem 4.10, the behavior of *remainder function* $\rho(\cdot)$ plays a significant role for the analysis of convergence. If a stronger condition is imposed on ρ (i.e., on the order of approximation for F), then Theorem 4.10 can be refined a little bit. The next corollary is in this sense.

Corollary 4.15 (local convergence revision). *Keep in mind all assumptions of Theorem 4.10, where* $\lambda^* = \limsup_{t\to 0} (t^{-1}\rho(t)) = 0$. *Then, the value* α *mentioned in Theorem 4.10 can be chosen such that the sequence* (x_k) *converges superlinearly to the solution* x^* *of* (4.13).

Proof. Let $\rho(t) = \rho_1(t)t$ for some real-valued function $\rho_1 : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ satisfying $\limsup_{t\to 0} \rho_1(t) = 0$. Then there exists a parameters $0 < \alpha < r/2$ such that

$$\begin{cases} \frac{3}{2}L\alpha + \varrho(\alpha) + \sup_{0\leqslant t\leqslant\alpha} \rho_1(t) < \frac{1}{\tau}, \forall t \leqslant \alpha, \\ \left[L\alpha + 3\varrho(\alpha) + 4\,|H(x^*)|^+ + \nu^*\right]\alpha + 2\left(\frac{1}{2}L\alpha^2 + \rho(\alpha)\right) < s. \end{cases} \tag{4.38}$$

Here, the positive constants τ, r, s and ν^* are selected as similar as the proof of Theorem 4.10.

Now, let x be in $\mathbb{B}(x^*, \alpha)$. Following the construction as in the proof of Theorem 4.10, we obtain sequence (x_k) induced by (4.14) for which $x_0 = x$ and, in addition,

$$\|x^* - x_{k+1}\| \leqslant \psi_1 (\|x^* - x_k\|) \|x^* - x_k\|, k = 0, 1, \ldots \tag{4.39}$$

Here, $\psi_1(t) := \frac{1}{1-[\tau Lt+\tau\varrho(t)]} \left[\frac{1}{2}\tau Lt + \tau\rho_1(t)\right]$, $t \in [0, \alpha]$. By induction, we can prove $\|x^* - x_k\| \leqslant \alpha$ and $0 \leqslant \sup_k \psi_1 (\|x^* - x_k\|) < 1$. Particularly, (4.39) shows that x_k converges to x^* as $k \to \infty$. Taking into account

$$\limsup_{t\to 0} \psi_1(t) = \limsup_{t\to 0} \left\{ \frac{1}{1 - [\tau Lt + \tau\varphi(t)]} \left[\frac{1}{2}\tau Lt + \tau\rho_1(t)\right]\right\} = 0,$$

the superlinear convergence is involved. This completes the proof of Corollary 4.15. □

4.3 A Numerical Illustration

We examine a simple example sketching the applicability of the convergence theorems in Section 4.2 before. Let's consider a cubic polynomial of real variable $f(x) = -(x-1)^3 + x - 1$. Choosing $X = Y = \mathbb{R}$ and

$$F(x) = \begin{cases} [\exp(-2x), +\infty), & \text{if } x \geqslant 0, \\ \emptyset, & \text{otherwise,} \end{cases} \tag{4.40}$$

where $\exp(\cdot)$ denotes the usual exponential function $\exp(x) = \sum_{n\geqslant 0} \frac{x^n}{n!}$. Figure 4.1 plots the graphs of $-f$ and F.

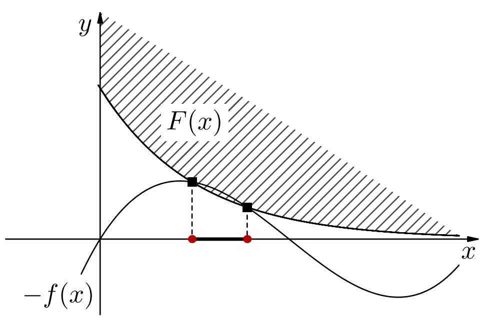


Fig. 4.1: The graphs of $-f(\cdot)$ and $F(\cdot)$ in numerical illustration

To apply the results proved in the previous section, we let

$$H(x)(u) = \begin{cases} \Big\{ -2u\exp(-2x) \Big\}, & \text{if } x \geqslant 0, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{4.41}$$

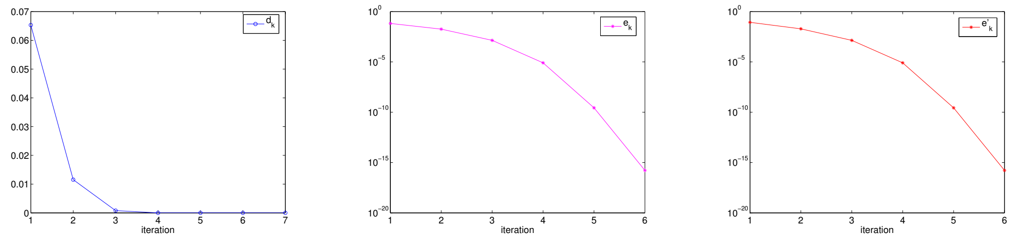
According to the definition in (4.17), it is possible to check that

$$\Lambda_H(x, x', t) = 2\left|\exp(-2x) - \exp(-2x')\right| t, \quad \text{for all } x, x' \geqslant 0 \text{ and } t \geqslant 0. \tag{4.42}$$

Furthermore, after some direct computations, under the following data

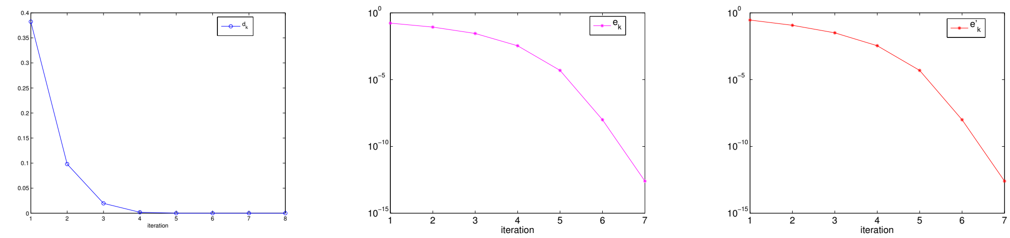
$$\rho(t) = \exp(2t) - 2t - 1, \varrho(t) = 2\left[\exp(2t) - 1\right], \tag{4.43}$$

the conditions (4.18) and (4.19) in Theorem 4.6 are fulfilled at any reference point $x > 0$. Similarly, the same conclusion is also valid for the assumptions (A1), (A2) and (A3) of Theorem 4.10 for any solution $x^* > 0$ (if exists) of the inclusion $0 \in f(x) + F(x)$. Some numerical performances are depicted by Figures 4.2 and 4.3.



(a): $d_k = d(0, f(x_k) + F(x_k))$ (b): $e_k = |x_{k+1} - x_k|$ (c): $e'_k = |x_k - x_{kmax}|$

Fig. 4.2: Numerical results: starting point $x_0 = 0.4$



(a): $d_k = d(0, f(x_k) + F(x_k))$ (b): $e_k = |x_{k+1} - x_k|$ (c): $e'_k = |x_k - x_{kmax}|$

Fig. 4.3: Numerical results: starting point $x_0 = 0.2$

Chapter 5

Newton-Type Algorithm in Riemannian Manifolds

As an overview, this chapter deals with the developments related to the inclusions of the form

$$0 \in \varphi(p) + \Phi(p). \tag{5.1}$$

Here, the variable p is in some manifold $\mathcal{M}$ of dimension m, $\varphi : \mathcal{M} \longrightarrow \mathbb{R}^n$ is a smooth map, and $\Phi : \mathcal{M} \rightrightarrows \mathbb{R}^n$ is a set-valued map. As in other chapters, we require that the graph $\operatorname{Gr} \Phi := \big\{(p, w) : w \in \Phi(p)\big\}$ is closed with respect to the product topology of $\mathcal{M} \times \mathbb{R}^n$.

For the aim of solving (5.1), we start at a guess point p_0 which is often expected to be nearby some proper solution, and generate an iterative sequence of approximation points. In details, suppose at k-step the iteration p_k is known, we choose a suitable retraction $R_k : T_{p_k}\mathcal{M} \longrightarrow \mathcal{M}$, and then update the succeeding term p_{k+1} through the subproblem

$$0 \in \varphi(p_k) + \mathscr{D}\varphi(p_k)(u_k) + (\Phi \circ R_k)(u_k), p_{k+1} = R_k(u_k). \tag{5.2}$$

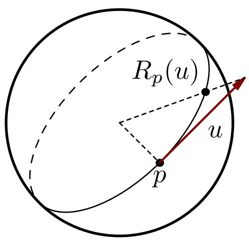


Fig. 5.1: Retraction on the sphere $\mathbb{S}^2$

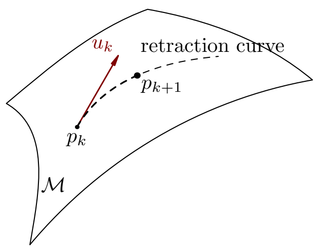


Fig. 5.2: Illustration of updating a new iteration

We notice that, when $\mathcal{M} = \mathbb{R}^m$, by setting $R_k(u) = p_k + u$ the usual translation, (5.2) recovers the Josephy-Newton method mentioned in Chapter 3. The concept of metric regularity property again plays a crucial role in order to analyze the convergence of algorithm (5.2). As a preparation, we begin by discussing the stability of metric regularity for mappings whose domains are in tangent spaces.

5.1 Some Preliminaries on Metric Regularity

Let p and q be two points in $\mathcal{M}$ such that $d_{\mathscr{R}}(p, q)$ is sufficiently small. This section will focus on the question how we can link the regularity property of two some mappings which are given on $T_p\mathcal{M}$ and $T_q\mathcal{M}$ respectively. For the purpose of studying the scheme (5.2), we may only work with a family of mappings $\mathscr{D}\varphi(z)(\cdot) + (\Phi \circ R_z)(\cdot)$, where R_z is a retraction at z. The following assumption will be essential and this will be kept throughout the whole of this chapter.

Standing Assumption. We suppose that all retractions R_z are well-defined on an open set Ω of $\mathcal{M}$ for which the next regularization condition holds. There is some subset $\Omega' \subset \Omega$ along with some real functions $\rho, \varrho, \delta : \Omega' \longrightarrow (0, +\infty)$ such that

$$\rho(z) \leqslant \frac{d_{\mathscr{R}}\big(R_z(v), R_z(v')\big)}{\|v - v'\|} \leqslant \varrho(z); \quad \text{if } z \in \Omega' \text{ and } v \neq v' \text{ in } \delta(z)\mathbb{B}_z. \tag{5.3}$$

Conventionally, we always fix the manifold $\mathcal{M}$ as well as an open (and convex) subset $\Omega \subset \mathcal{M}$. In addition, we simply write $R \in \mathbf{URC}\big(\rho, \varrho, \delta, \Omega'\big)$ to indicate the property described by (5.3).

Such a restriction like (5.3) is fulfilled in a wide range of practical applications. It is easy to see that when $\mathcal{M} = \mathbb{R}^m$ and $R_z(v) = z + v$, (5.3) holds itself without setting more, just taking $\Omega' = \Omega = \mathbb{R}^m$, $\rho(\cdot) = \varrho(\cdot) \equiv 1$. Another simple case being less of triviality is shown in Example 5.1 below. On any Riemannian variety with geodesic retraction $R = \mathrm{Exp}$, it also holds due to the local property of exponential map (cf. [16, 41, 55]). Otherwise, a criterion ensures the validity of (5.3) was suggested in [68], where the collection of retractions R_z are assumed to have equicontinuous derivatives.

Example 5.1 (unit sphere). Let $\mathbb{S}^{m-1} = \big\{p \in \mathbb{R}^m : p^T p = 1\big\}$ be the unit sphere in $\mathbb{R}^m$. $\mathbb{S}^{m-1}$ is endowed with the Riemannian metric inherited from Euclidean distance on $\mathbb{R}^m$ [1]. For each $p \in \mathbb{S}^{m-1}$, the tangent space $T_p\mathbb{S}^{m-1}$ can be identified to $p^\perp := \big\{u \in \mathbb{R}^m : u^T p = 0\big\}$. Consider the retraction given by $R_p(u) = \frac{p+u}{\|p+u\|}$ for $u \in p^\perp$ and $\|\cdot\|$ is Euclidean norm. The case $m = 3$ was depicted in Figure 5.1.

We are able to check that $d_{\mathcal{R}}(p,q) = \arccos\big(p^T q\big)$, which yields $d_{\mathcal{R}}\big(p, R_p(u)\big) = \arctan\|u\|$. As a result, R_p is injective in the whole tangent space $T_p\mathbb{S}^{m-1}$.
Let $q \in \mathbb{S}^{m-1}$ and $v, v' \in T_q\mathbb{S}^{m-1}$. Abbreviating $\hat{d} = d_R\big(R_q(v), R_q(v')\big)$, then $\cos\hat{d} = \frac{(q+v)^T(q+v')}{\|q+v\|\|q+v'\|}$, which is equivalent to $\cos\hat{d} = \frac{1+v^Tv'}{\sqrt{(1+\|v\|^2)(1+\|v'\|^2)}}$. Thus,

$$\sin^2\hat{d} = \frac{\|v-v'\|^2 + \|v\|^2\|v'\|^2 - \big(v^Tv'\big)^2}{(1+\|v\|^2)(1+\|v'\|^2)}. \tag{5.4}$$

Using the inner inequality $v^Tv' \leqslant \|v\|\,\|v'\|$, (5.4) implies

$$\sin^2\hat{d} \geqslant \frac{\|v-v'\|^2}{(1+\|v\|^2)(1+\|v'\|^2)}. \tag{5.5}$$

On the other hand, thanks to the equality $2v^Tv' = \|v\|^2 + \|v'\|^2 - \|v-v'\|^2$, we get

$$\begin{aligned}\|v\|^2\|v'\|^2 - \big(v^Tv'\big)^2 &= \|v\|^2\|v'\|^2 - \frac{1}{4}\Big(\|v\|^2+\|v'\|^2-\|v-v'\|^2\Big)^2\\ &= \frac{1}{2}\Big(\|v\|^2+\|v'\|^2\Big)\|v-v'\|^2 - \frac{1}{4}\Big[\|v-v'\|^4 + \Big(\|v\|^2-\|v'\|^2\Big)^2\Big]\\ &\leqslant \frac{1}{2}\Big(\|v\|^2+\|v'\|^2\Big)\|v-v'\|^2.\end{aligned}$$

Consequently,

$$\sin^2\hat{d} \leqslant \frac{1+\frac{1}{2}\left(\|v\|^2+\|v'\|^2\right)}{\left(1+\|v\|^2\right)\left(1+\|v'\|^2\right)}\|v-v'\|^2 \leqslant \|v-v'\|^2. \tag{5.6}$$

Observe that one has $\frac{2}{\pi} \leqslant \frac{\sin t}{t} \leqslant 1$ as long as $0 < t \leqslant \frac{2}{\pi}$. Based on (5.5) and (5.6), we deduce

$$\frac{1}{1+r^2}\|v-v'\| \leqslant d_{\mathcal{R}}\big(R_q(v), R_q(v')\big) \leqslant \frac{\pi}{2}\|v-v'\|$$

whenever $\|v\| \leqslant r$ and $\|v'\| \leqslant r$, with $0 < r \leqslant 1$. Fix $r \in (0,1)$. Under the substitution $\delta(z) = r$, $\rho(z) = \frac{1}{1+r^2}$ and $\varrho(z) = \frac{\pi}{2}$, we find $R \in \mathbf{URC}\big(\rho, \varrho, \delta, \mathbb{S}^{m-1}\big)$.

Stability results for a class of multifunctions of the form $\mathscr{D}\varphi(z)(\cdot) + (\Phi \circ R_z)(\cdot)$ were recently established in [4, Propositions 3.1] and [4, Propositions 3.3]. They are based on the suppositions that both functions ρ and δ are bounded from below whereas $\sup_{z\in\Omega}\varrho(z) < +\infty$. The next Propositions 5.2 and 5.5 are going to provide the refinements of those aforementioned results. However, as we will see in Section 5.2, when applying these propositions for the study of algorithm (5.2), it seems necessarily to impose some conditions on ρ, ϱ and δ used in [4].

Proposition 5.2 (local stability). *Let $\{R_z : z \in \Omega\}$ be a given family of retractions and (λ_R, ι_R) be a normal pair for Ω. Fix a point $p \in \Omega$ and suppose $R \in \mathbf{URC}(\rho, \varrho, \delta, W_p)$ where $W_p = \mathbb{B}_{\mathcal{M}}(p, \lambda_R(p))$. Pick a tangent vector $\bar{u} \in T_p\mathcal{M}$ such that $\kappa = \operatorname{Reg}\Psi_p(\bar{u}, \bar{x}) < +\infty$ with $\Psi_p(\cdot) = \mathscr{D}\varphi(p)(\cdot) + (\Phi R_p)(\cdot)$ and $\bar{x} \in \Psi_p(\bar{u})$. Consider some positive constants α, β, $\sigma > 0$, ρ_0 and ϱ_0 fulfilling*

$$\theta = \varrho_0 \kappa \left(\alpha + \beta \left\|\mathscr{D}\varphi(p)\right\|\right) < \rho_0, \varrho_0 \left\|\bar{u}\right\|_p + \sigma < \rho_0 \iota_R(p), \rho_0 \leqslant \rho(p) \leqslant \varrho(p) \leqslant \varrho_0.$$

Choose a point $q \in W_p$ with $d_{\mathcal{R}}(p,q) \leqslant \sigma$ so that the conditions (i) *and* (ii) *below are satisfied for each geodesic path $\chi : [a,b] \longrightarrow W_p$ connecting $\chi(a) = p$ to $\chi(b) = q$:*

(i) $\Sigma_{\chi,q,p} = R_p^{-1} R_q - P_\chi^{b,a}$ is Lipschitz continuous on the ball $\iota_R(p)\mathbb{B}_q$ with modulus β;

(ii) $\|G_{\chi,p,q}\| \leqslant \alpha$, where $G_{\chi,p,q} := \mathscr{D}\varphi(q) P_\chi^{a,b} - \mathscr{D}\varphi(p)$ is a linear map.

If $\iota_R(p) \leqslant \min\{\delta(p), \delta(q)\}$, $\rho_0 \leqslant \rho(q) \leqslant \varrho(q) \leqslant \varrho_0$ and $\varrho_0 \left\|\bar{u}\right\|_p < \lambda_R(p)$, then we obtain $\operatorname{Reg}\Psi_q(\bar{v}, \bar{y}) \leqslant \frac{1}{1-\theta}\kappa$. Here, $\bar{v}$ is a tangent vector in $\iota_R(p)\mathbb{B}_q$ such that $R_q(\bar{v}) = R_p(\bar{u})$, $\bar{y} = \bar{x} - \mathscr{D}\varphi(p)(\bar{u}) + \mathscr{D}\varphi(q)(\bar{v})$ whereas $\Psi_q(\cdot) = \mathscr{D}\varphi(q)(\cdot) + (\Phi R_q)(\cdot)$.

Before proving the preceding statement, we recall first all norms of linear operators in Proposition 5.2 are taken with respect to the scalar products of the corresponding spaces. Lemma 5.3 below will be useful for the proof of Proposition 5.2.

Lemma 5.3. *Define $\Lambda_{q,p} := \mathscr{D}\varphi(q) - \mathscr{D}\varphi(p) R_p^{-1} R_q$. If p can be linked to q by a geodesic segment which totally lies inside W_p, then for each $r \leqslant \iota_R(p)$ the map $\Lambda_{q,p}$ is Lipschitz continuous on $r\mathbb{B}_q$ with a modulus $L_q = \|\mathscr{D}\varphi(p)\| \beta + \alpha$.*

Proof. Let χ be a geodesic such that $\chi(0) = p$, $\chi(1) = q$ and $\chi([0,1]) \subset W_p$. Then, we have the expression $\Lambda_{q,p} = -\mathscr{D}\varphi(p)\Sigma_{\chi,q,p} + G_{\chi,p,q} P_\chi^{1,0}$. Recall that $P_\chi^{1,0}$ has unit norm, so the conclusion of this lemma follows by using simultaneously properties (i) and (ii). □

Proof of Proposition 5.2. Without lost of generality, we can assume W_p to be a convex neighborhood of $\mathcal{M}$ at p [20]. For simplicity, we denote $\mu_p = \|\mathscr{D}\varphi(p)\|$, $\lambda_0 = \lambda_R(p)$, $\eta_0 = \varrho_0/\rho_0$, $\iota_0 = \iota_R(p)$, and $\delta_0 = \min\{\delta(p), \delta(q)\}$. Alternatively, because the spaces are specified in the context, we use the common notations $\|\cdot\|$

and $d(\cdot,\cdot)$ respectively for any norm and distance function. Pick some parameters $\kappa' \geqslant \kappa$, $r > 0$ and $s > 0$ so that Ψ_p is metrically regular with respect to a modulus κ' on the neighborhood

$$\mathcal{V} = \big\{(u,x) : \|u - \bar{u}\| < r, \|x - \bar{x}\| < s\big\}$$

and that $\theta' = \kappa' (\alpha + \beta\mu_p)\, \eta_0 < 1$. Let us now take $r' > 0$ and $s' > 0$ for which the group of four coming inequalities (5.7a), (5.7b), (5.7c) and (5.7d) is valid as well

$$\eta_0 \frac{2}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s' < r, \tag{5.7a}$$

$$(\eta_0)^2 (\alpha + \beta\mu_p) \frac{2}{1-\theta'} r' + \frac{1}{1-\theta'} s' < s, \tag{5.7b}$$

$$\eta_0 \|\bar{u}\| + \sigma/\rho_0 + \eta_0 \frac{2}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s' < \iota_0, \tag{5.7c}$$

$$\varrho_0 \|\bar{u}\| + \varrho_0 \left(\eta_0 \frac{2}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s'\right) < \lambda_0. \tag{5.7d}$$

We are going to show that Ψ_q is metrically regular with modulus $\tau' = \eta_0 \frac{1}{1-\theta'} \kappa'$ on the neighborhood $\mathcal{V}' = \big\{(v,y) : \|v - \bar{v}\| < r', \|y - \bar{y}\| < s'\big\}$. Indeed, fix (v,y) in $\mathcal{V}'$, and write $C = d\big(y, \Psi_q(v)\big)$. By setting $v_0 = v$, one has $\|v_0\| \leqslant \|v_0 - \bar{v}\| + \|\bar{v}\|$. Recalling $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$ and $\|\bar{v}\| < \iota_0 \leqslant \delta(q)$, we get

$$\begin{aligned}
\|\bar{v}\| &\leqslant \rho(q)^{-1} d\big(q, R_q(\bar{v})\big) \leqslant \rho(q)^{-1} \left[d(q,p) + d\big(p, R_p(\bar{u})\big)\right] \\
&\leqslant \rho(q)^{-1} \left[\sigma + \varrho(p) \|\bar{u}\|\right] \\
&\leqslant \sigma/\rho_0 + \eta_0 \|\bar{u}\| .
\end{aligned}$$

Thus, the value of quantity $\|v_0\|$ can be majorized as follows

$$\begin{aligned}
\|v_0\| &\leqslant \|v_0 - \bar{v}\| + \rho(q)^{-1}\sigma + \rho(q)^{-1}\varrho(p) \|\bar{u}\| \\
&< r' + \rho(q)^{-1}\sigma + \rho(q)^{-1}\varrho(p) \|\bar{u}\| \\
&< \iota_0 \leqslant \min\{\delta(p), \delta(q)\}.
\end{aligned}$$

Denoting $z_0 = R_q(v_0) \in W_p$, $\bar{z} = R_q(\bar{v}) = R_p(\bar{u})$. In terms of normal pair, there exists a unique $u_0 \in \iota_0 \mathbb{B}_p$ with $R_p(u_0) = R_q(v_0)$. Since $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$ and $\iota_0 \leqslant \delta(p)$, we deduce $d\big(z_0, \bar{z}\big) \leqslant \varrho(q) \|v_0 - \bar{v}\|$ and $\|u_0 - \bar{u}\| \leqslant \rho(p)^{-1} d(z_0, \bar{z})$. Particularly, these arguments give us

$$\|u_0 - \bar{u}\| \leqslant d(z_0, \bar{z})/\rho_0 \leqslant \varrho_0 \|v_0 - \bar{v}\| / \rho_0 = \eta_0 \|v_0 - \bar{v}\| . \tag{5.8}$$

According to (5.7a), it holds that $\|u_0 - \bar{u}\| < r$. Define $x_0 = y + \mathscr{D}f(p)(u_0) - \mathscr{D}f(v_0)$, then we have

$$\|x_0 - \bar{x}\| \leqslant \|y - \bar{y}\| + \|\mathscr{D}f(p)(u_0) - \mathscr{D}f(v_0) - [\mathscr{D}f(p)(\bar{u}) - \mathscr{D}f(\bar{v})]\|.$$

Remind $R_p(u_0) = R_q(v_0)$ and $R_p(\bar{u}) = R_q(\bar{v})$, Lemma 5.3 can be applied. As a result, we obtain

$$\begin{aligned} \|x_0 - \bar{x}\| &\leqslant \|y - \bar{y}\| + \|\Lambda_{q,p}(v_0) - \Lambda_{q,p}(\bar{v})\| < s' + (\alpha + \beta\mu_p)\|v_0 - \bar{v}\| \\ &< s' + (\alpha + \beta\mu_p) r'. \end{aligned} \tag{5.9}$$

Inasmuch as $\eta_0 \geqslant 1$, (5.7b) and (5.9) show that $\|x_0 - \bar{x}\| < s$, which implies $(u_0, x_0) \in \mathcal{V}$. By invoking $\kappa' \in \operatorname{Regmod}\big(\Psi_p, \mathcal{V}\big)$, we deduce

$$d\big(u_0, \Psi_p^{-1}(x_0)\big) \leqslant \kappa' d\big(x_0, \Psi_p(u_0)\big).$$

Let's choose a tangent vector $u_1 \in \Psi_p^{-1}(x_0)$ such that $\|u_0 - u_1\| = d\big(u_0, \Psi_p^{-1}(x_0)\big)$. Then, the latter yields

$$\begin{aligned} \|u_0 - u_1\| &\leqslant \kappa' d\big(x_0, \Psi_p(u_0)\big) = d\big(x_0, \mathscr{D}\varphi(p)(u_0) + (\Phi R_p)(u_0)\big) \\ &= \kappa' d\big(y - \mathscr{D}\varphi(q)(v_0), (\Phi R_q)(v_0)\big) = \kappa' C. \end{aligned}$$

To continue, we set $z_1 = R_p(u_1)$ and claim $z_1 \in W_p$. Thanks to the triangle inequality in $T_p\mathcal{M}$, we find

$$\|u_0 - u_1\| = d\big(u_0, \Psi_p^{-1}(x_0)\big) \leqslant \|u_0 - \bar{u}\| + d\big(\bar{u}, \Psi_p^{-1}(x_0)\big).$$

Observe that the pair $(\bar{u}, x_0)$ is in $\mathcal{V}$. Consequently, $d\big(\bar{u}, \Psi_p^{-1}(x_0)\big) \leqslant \kappa' d\big(x_0, \Psi_p(\bar{u})\big)$. Because of $\bar{x} \in \Psi_p(\bar{u})$, it is possible to write

$$\|u_0 - u_1\| \leqslant \|u_0 - \bar{u}\| + \kappa' \|x_0 - \bar{x}\|. \tag{5.10}$$

In combination with (5.8) and (5.9), (5.10) gives us

$$\|u_0 - u_1\| < \eta_0 r' + \kappa' [s' + (\alpha + \beta\mu_p) r'] \leqslant \eta_0 (1 + \theta') r' + \kappa' s'.$$

Hence, the following estimation is valid

$$\|u_1 - \bar{u}\| \leqslant \|u_0 - \bar{u}\| + \|u_0 - u_1\| \leqslant \eta_0 \|v_0 - \bar{v}\| + \|u_0 - u_1\|$$

$$< \eta_0 \,(2+\theta')\, r' + \kappa' s' < r.$$

In the space $T_p\mathcal{M}$, one has $\|u_1\| \leqslant \|\bar{u}\| + \|u_0 - u_1\|$, which provides

$$\|u_1\| < \|\bar{u}\| + \eta_0 \,(2+\theta')\, r' + \kappa' s' < \iota^* \leqslant \delta(p).$$

Because of $R \in \mathbf{URC}\,(\rho, \varrho, \delta, W_p)$, $z_1 = R_p(u_1)$ belongs to W_p. Assigning $v_1 = R_q^{-1}(z_1) \in \iota^*\mathbb{B}_q$, and $x_1 = y + \mathscr{D}\varphi(p)(u_1) - \mathscr{D}\varphi(q)(v_1)$, we go to the next step of finding u_2, v_2 and so on.

We proceed to the inductive process. Suppose the tangent vectors $u_0, \ldots, u_k \in \iota^*\mathbb{B}_p$ and $v_0, \ldots, v_k \in \iota^*\mathbb{B}_q$ are known. Additionally, as suggested from above, we assume those elements admit the following relations:

- $R_p(u_j) = R_q(v_j)$;
- $u_{j+1} \in \Psi_p^{-1}(x_j)$ for $x_j = y + \mathscr{D}\varphi(p)(u_j) - \mathscr{D}\varphi(q)(v_j)$;
- $\|u_j - u_{j+1}\| \leqslant (\theta')^j \,\|u_0 - u_1\|$.

For the goal of generating u_{k+1} and v_{k+1}, let's consider the pair (u_k, x_k), where $x_k = y + \mathscr{D}\varphi(p)(u_k) - \mathscr{D}\varphi(q)(v_k)$. Involving again the triangle inequality in $T_p\mathcal{M}$

$$\|u_k - \bar{u}\| \leqslant \sum_{j=0}^{k-1} \|u_j - u_{j+1}\| + \|u_0 - \bar{u}\| \leqslant \sum_{j=0}^{k-1} (\theta')^j \,\|u_0 - u_1\| + \|u_0 - \bar{u}\|.$$

In view of (5.8), it follows from the estimation $\|u_0 - u_1\| < \eta_0 \,(1+\theta')\, r' + \kappa' s'$ that

$$\|u_k - \bar{u}\| < \frac{1}{1-\theta'} \left[\eta_0 \,(1+\theta')\, r' + \kappa' s'\right] + \eta_0 r' = \eta_0 \frac{2}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s' < r.$$

Otherwise, thanks to $x_k = \Lambda_{q,p}(v_k)$, Lemma 5.3 yields

$$\|x_k - \bar{x}\| \leqslant (\alpha + \beta\mu_p)\,\|v_k - \bar{v}\|.$$

Recalling $R \in \big(\rho, \varrho, \delta, W_p\big)$, it infers from the facts $u_k \in \iota^*\mathbb{B}_p$ and $v_k \in \iota^*\mathbb{B}_q$ that

$$\begin{aligned}\|v_k - \bar{v}\| \leqslant \frac{1}{\rho_0} d\big(R_q(v_k), R_q(\bar{v})\big) = \frac{1}{\rho_0} d\big(R_p(u_k), R_p(\bar{u})\big) \leqslant \eta_0 \,\|u_k - \bar{u}\| \\ < \eta_0 \left[\eta_0 \frac{2}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s'\right].\end{aligned}$$

Consequently, (u_k, x_k) belongs to $\mathcal{V}$, since

$$\|x_k - \bar{x}\| < (\eta_0)^2 (\alpha + \beta\mu_p) \frac{2}{1-\theta'} r' + \eta_0 \frac{\theta'}{1-\theta'} s' < s.$$

Invoking the fact $\kappa' \in \operatorname{Regmod}\big(\Psi_p, \mathcal{V}\big)$ once more, we find

$$d\big(u_k, \Psi_p^{-1}(x_k)\big) \leqslant \kappa' d\big(x_k, \Psi_p(u_k)\big) \leqslant \kappa' \|x_k - x_{k-1}\|.$$

The set $\Psi_p^{-1}(x_k)$ is closed and nonempty in $T_p\mathcal{M}$, so it contains an element, written as u_{k+1}, such that $\|u_k - u_{k+1}\| = d\big(u_k, \Psi_p^{-1}(x_k)\big)$. Repeating the same arguments as the preceding cases of $\|x_k - \bar{x}\|$ and $\|v_k - \bar{v}\|$, it is possible to prove that

$$\begin{cases} \|x_k - x_{k-1}\| \leqslant (\alpha + \beta\mu_p) \|v_k - v_{k-1}\|, \\ \|v_k - v_{k-1}\| \leqslant \eta_0 \|u_k - u_{k-1}\|. \end{cases}$$

In summary, we derive from these arguments

$$\|u_k - u_{k+1}\| \leqslant \kappa'(\alpha + \beta\mu_p)\eta_0 \|u_k - u_{k-1}\| \leqslant (\theta')^k \|u_0 - u_1\| \leqslant (\theta')^k \kappa' C.$$

To see that $u_{k+1} \in \iota^*\mathbb{B}_p$, we estimate as follows

$$\begin{aligned} \|u_{k+1}\| &\leqslant \|\bar{u}\| + \sum_{j=0}^{k} \|u_j - u_{j+1}\| \leqslant \|\bar{u}\| + \sum_{j=0}^{k} (\theta')^j \|u_0 - u_1\| \\ &< \|\bar{u}\| + \frac{1}{1-\theta'} \left[\eta_0 (1+\theta') r' + \kappa' s'\right] \leqslant \|\bar{u}\| + \frac{2\eta_0}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s' \\ &< \iota^*. \end{aligned}$$

Thus, by virtue of $\iota^* \leqslant \delta(p)$, it holds that

$$d\big(R_p(u_{k+1}), p\big) \leqslant \varrho_0 \|u_{k+1}\| \leqslant \varrho_0 \left(\|\bar{u}\| + \frac{2\eta_0}{1-\theta'} r' + \frac{\kappa'}{1-\theta'} s' \right) < \lambda^*,$$

where the last inequality is due to (5.7d). This means, the point $z_{k+1} = R_p(u_{k+1})$ lies into W_p. As a result, it has a unique tangent vector $v_{k+1} \in \iota^*\mathbb{B}_q$ satisfying $R_q(v_{k+1}) = z_{k+1}$. The sequences (u_k) and (v_k) are completely determined.

According to the construction, one has $\|u_k - u_{k+1}\| \leqslant (\theta')^k \kappa' C$. Since $\theta' < 1$, the positive series $\sum_{j\geqslant 0} (\theta')^j \kappa' C$ is convergent. Therefore, u_k converges in $T_p\mathcal{M}$ to some u^*. Furthermore, we also infer $\|v_k - v_{k-1}\| \leqslant \eta_0 \|u_k - u_{k-1}\|$ from the construction above, so the limit $\lim\limits_{k\to\infty} v_k = v^*$ exists in $T_q\mathcal{M}$.

Remind $R_q(v_k) = R_p(u_k)$ and note that all smooth maps are continuous, we arrive $R_q(v^*) = R_p(u^*)$. Taking into account $y + \mathscr{D}\varphi(p)(u_k) - \mathscr{D}\varphi(q)(v_k) \in \Psi_p(u_{k-1})$, we obtain $y + \mathscr{D}\varphi(p)(u^*) - \mathscr{D}\varphi(q)(v^*) \in \Psi_p(u^*)$ after passing to the limit in k. Equivalently, this means

$$y \in \mathscr{D}\varphi(q)(v^*) + \Psi_p(u^*) = \mathscr{D}\varphi(q)(v^*) + (\Phi R_q)(v^*) = \Psi_q(v^*).$$

Hence, $d\big(v, \Psi_q^{-1}(y)\big) \leqslant \|v - v^*\| = \|v_0 - v^*\|$. However, by using the triangle inequality and the relation $\|v_k - v_{k-1}\| \leqslant \eta_0 \|u_k - u_{k-1}\|$, we obtain

$$\begin{aligned} \|v_0 - v^*\| &\leqslant \sum_{k\geqslant 0} \|v_k - v_{k+1}\| \leqslant \sum_{k\geqslant 0} \eta_0 \|u_k - u_{k+1}\| \leqslant \eta_0 \sum_{k\geqslant 0} (\theta')^k \kappa' C \\ &= \eta_0 \frac{\kappa'}{1-\theta'} C. \end{aligned} \tag{5.11}$$

Thus, Ψ_q is metrically regular on $\mathcal{V}'$ with a modulus $\tau' = \eta_0 \frac{\kappa'}{1-\theta'}$. Since τ' can be arbitrarily close to τ, we reach to the conclusion $\operatorname{Reg} \Psi_q(\bar{v}, \bar{y}) \leqslant \tau$. □

Remark 5.4. By adding $\inf_{z\in W_p} \rho(z) > 0$, $\inf_{z\in W_p} \delta(z) > 0$ and $\sup_{z\in W_p} \varrho(z) < +\infty$, then Proposition 5.2 subsumes to the corresponding one proved in [4]. Indeed, we have only to replace $\iota_R^*(z) := \min\big\{\iota_R(z), \inf_{z\in W_p} \delta(z)\big\}$ with $\iota_R(z)$ if necessary. Although these restrictions seem to be more than enough for the validity of Proposition 5.2, we shall need them when analyzing the behaviour of algorithm (5.2) in the next section. The reason is that, in order to get the succeeding point p_{k+1} from the current step p_k, we have to invoke the metric regularity property for $\mathscr{D}f(p_k)(\cdot) + (\Phi R_{p_k})(\cdot)$ (which assume that such a property should be stable over elements of approximating sequence).

Proposition 5.5 (semi-local stability)**.** *Similarly as Proposition 5.2, we fix $p \in \Omega \subset \mathcal{M}$ (Ω is open) and a normal pair (λ_R, ι_R) associated with a given retraction R. Keep in mind the assumption $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$, where $W_p = \mathbb{B}_{\mathcal{M}}(p, \lambda_R(p))$ is also assumed to be a convex neighborhood at p. Suppose that the mapping $\Psi_p := \mathscr{D}\varphi(p) + (\Phi R_p)$ is metrically regular on the set*

$$V_{r,s}(\Psi_p) := \big\{(u, x) \in T_p\mathcal{M} \times \mathbb{R}^n : \|u\| \leqslant r, d\big(x, \Psi_p(u)\big) \leqslant s\big\}$$

together with a modulus $\kappa > 0$. Consider a point $q \in W_p$ with $d_{\mathcal{R}}(p, q) \leqslant \sigma$ for some $\sigma > 0$. Let s, r', $s' \leqslant s$, α and β be positive real numbers which adopt the *following conditions*

$$\theta = \varrho_0 \kappa(\alpha + \beta\mu_p)/\rho_0 < 1, \tag{5.12a}$$

$$\sigma/\rho_0 + \varrho_0 r'/\rho_0 + \frac{\kappa}{1-\theta} s' < \min\left\{\iota_R(p), \delta(p), \delta(q), r\right\}, \tag{5.12b}$$

$$(\varrho_0)^2 r'/\rho_0 + \varrho_0 \sigma/\rho_0 + \varrho_0 \frac{\kappa}{1-\theta} s' < \lambda_R(p), \tag{5.12c}$$

where $\mu_p = \|\mathscr{D}\varphi(p)\|$, $\rho_0 = \min\{\rho(p), \rho(q)\}$ *and* $\varrho_0 = \max\{\varrho(p), \varrho(q)\}$. *In addition, we require the both suppositions* (i) *and* (ii) *in Proposition 5.2 hold with respect to each geodesic segment* $\chi : [a, b] \longrightarrow W_p$ *having* $\chi(a) = p$, $\chi(b) = q$. *Then* $\Psi_q := \mathscr{D}\varphi(q) + (\Phi R_q)$ *is metrically regular with modulus* $\tau = \frac{\varrho_0}{\rho_0}\frac{1}{1-\theta}\kappa$ *on the set*

$$V_{r',s'}(\Psi_q) := \left\{(v, y) \in T_q\mathcal{M} \times \mathbb{R}^n : \|v\| \leqslant r', d\big(y, \Psi_q(v)\big) \leqslant s'\right\}.$$

Proof. Keeping the notations $\iota_0 = \iota_R(p)$, $\lambda_0 = \lambda_R(p)$, $\delta_0 = \min\{\delta(p), \delta(q)\}$ and $\eta_0 = \varrho_0/\rho_0$. Interchanging $\iota_R^*(z) = \min\{\iota_R(z), \delta(p), \delta(q)\}$ with $\iota_R(z)$ if necessary, we can assume $\iota_0 = \min\{\iota_R(p), \delta(p), \delta(q)\}$. Let's fix $(v, y) \in V_{r',s'}(\Psi_q)$ with $y \notin \Psi_q(v)$. The strategy of this proof is similar to the one of Proposition 5.2. At first, we set $C = d\big(y, \Psi_q(v)\big) > 0$, $v_0 = v$ and $z_0 = R_q(v_0)$. From (5.12b), we have $\|v_0\| \leqslant r' < \delta(q)$. Thus, the fact $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$ gives us

$$d(z_0, q) = d\big(R_q(v_0), q\big) \leqslant \varrho(q)\,\|v_0\| \leqslant \varrho_0 r'.$$

Observing $\eta_0 = \varrho_0/\rho_0 \geqslant 1$, it follows from (5.12c) that

$$d(z_0, p) \leqslant d(z_0, q) + d(q, p) \leqslant \varrho_0 r' + \sigma < \lambda_R(p).$$

Hence, z_0 belongs to W_p. In terms of normal pair, there exists tangent vector $u_0 \in \iota_0\mathbb{B}_p$ satisfying $R_p(u_0) = R_q(v_0)$. Put $x_0 = y + \mathscr{D}\varphi(p)(u_0) - \mathscr{D}\varphi(q)(v_0)$, one has

$$\begin{aligned} d\big(x_0, \Psi_p(u_0)\big) &= d\big(y - \mathscr{D}\varphi(q)(v_0), (\Phi R_p)(u_0)\big) \\ &= d\big(y, \mathscr{D}\varphi(q)(v_0) + (\Phi R_q)(u_0)\big) \\ &= C \leqslant s' \leqslant s. \end{aligned}$$

Moreover, the fact $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$ provides

$$\|u_0\| \leqslant \frac{1}{\rho_0} d\big(z_0, p\big) \leqslant \eta_0 r' + \frac{1}{\rho_0}\sigma < r.$$

Therefore, $(u_0, x_0) \in V_{r,s}(\Psi_p)$. Invoking the metric regularity property for Ψ_p, we can find an element $u_1 \in \Psi_p^{-1}(x_0)$ such that $\|u_0 - u_1\| \leqslant \kappa d\big(x_0, \Psi_p(u_0)\big)$. But $d\big(x_0, \Psi_p(u_0)\big) = C$, so we obtain $\|u_0 - u_1\| \leqslant \kappa C$.

Noticing $\kappa C \leqslant \kappa s' < \frac{\kappa}{1-\theta} s'$ and $\|u_0\| \leqslant \eta_0 r' + \frac{1}{\rho_0}\sigma$, the supposition (5.12b) implies $\|u_1\| \leqslant \eta_0 r' + \frac{1}{\rho_0}\sigma + \kappa s' < \iota_0$. Define a new point $z_1 = R_p(u_1)$, we get

$$d\big(z_1, p\big) \leqslant \varrho(p)\, \|u_1\| \leqslant \varrho_0 \left(\eta_0 r' + \frac{1}{\rho_0}\sigma + \kappa s'\right) < \lambda_0.$$

In other words, z_1 is in W_p. As a result, there exists a tangent vector $v_0 \in \iota^* \mathbb{B}_q$ satisfying $R_q(v_1) = z$.

Passing to the induction step, let $u_0, \ldots, u_k \in \iota^* \mathbb{B}_p$ and $v_0, \ldots, v_k \in \iota^* \mathbb{B}_q$ be given tangent vectors. Furthermore, based on the preceding arguments, it should be required that those vectors obey the constraints below:

- $R_p(u_i) = R_q(v_i)$;
- $u_{i+1} \in \Psi_p^{-1}(x_i)$ for $x_i = y + \mathscr{D}\varphi(p)(u_i) - \mathscr{D}\varphi(q)(v_i)$;
- $\|u_i - u_{i+1}\| \leqslant \theta^i \kappa C$.

Towards the aim of generating $u_{k+1} \in T_p\mathcal{M}$ and $v_{k+1} \in T_q\mathcal{M}$, we set $x_k = y + \mathscr{D}\varphi(p)(u_k) - \mathscr{D}\varphi(q)(v_k)$ and consider the pair (u_k, x_k) in $T_p\mathcal{M} \times \mathbb{R}^n$. Thanks to the triangle inequality, we have

$$\|u_k\| \leqslant \|u_0\| + \sum_{i=0}^{k-1} \|u_i - u_{i+1}\| \leqslant r' + \sum_{i=0}^{k-1} \theta^i \kappa C \leqslant r' + \frac{1}{1-\theta}\kappa s' < \iota_0.$$

By virtue of $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$, we deduce

$$d\big(R_p(u_k), p\big) \leqslant \varrho(p)\, \|u_k\| \leqslant \varrho_0 r' + \varrho_0 \frac{1}{1-\theta}\kappa s' < r.$$

Recall that $u_k \in \Psi_p^{-1}(x_{k-1})$, we infer $d\big(x_k, \Psi_p(u_k)\big) \leqslant \|x_k - x_{k-1}\|$. Following the same arguments as in the proof of Proposition 5.2, we can prove

$$\|x_k - x_{k-1}\| = \|\Lambda_{p,q}(v_k) - \Lambda_{p,q}(v_{k-1})\| \leqslant (\alpha + \beta\mu_p)\, \|v_k - v_{k-1}\|.$$

Since $R_p(u_j) = R_q(v_j)$, assumption $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$ allows us to write

$$\begin{aligned}\|v_k - v_{k-1}\| &\leqslant \frac{1}{\rho(q)} d\big(R_q(v_k), R_q(v_{k-1})\big) \leqslant \frac{1}{\rho_0} d\big(R_p(u_k), R_p(u_{k-1})\big) \\ &\leqslant \eta_0\, \|u_k - u_{k-1}\|.\end{aligned}$$

Consequently, involving the condition $\|u_k - u_{k-1}\| \leqslant \theta^{k-1}\kappa C$, one has

$$d\big(x_k, \Psi_p(u_k)\big) \leqslant \|x_k - x_{k-1}\| \leqslant \eta_0(\alpha + \beta\mu_p)\,\|u_k - u_{k-1}\| \\ \leqslant \eta_0(\alpha + \beta\mu_p)\theta^{k-1}\kappa C.$$

Taking into account $C \leqslant s' \leqslant s$ and $\eta_0(\alpha+\beta\mu_p)\kappa = \theta < 1$, (u_k, x_k) is in $V_{r,s}(\Psi_p)$. Hence, the metric regularity property of Ψ_p applied to (u_k, x_k) guarantees the existence of a vector $u_{k+1} \in \Psi_p^{-1}(x_k)$ such that $\|u_k - u_{k+1}\| \leqslant \kappa d\big(x_k, \Psi_p(u_k)\big)$. In particular, $\|u_k - u_{k+1}\| \leqslant \theta^k \kappa C$.

Since $\sum_{i=0}^{k} \theta^k\kappa C \leqslant \frac{1}{1-\theta}\kappa C \leqslant \frac{1}{1-\theta}\kappa s'$, triangle inequality in $T_p\mathcal{M}$ gives us

$$\|u_{k+1}\| \leqslant \|u_0\| + \sum_{i=0}^{k} \|u_i - u_{i+1}\| \leqslant \eta_0 r' + \frac{1}{\rho_0}\sigma + \frac{1}{1-\theta}\kappa s' < \iota_0.$$

According to the fact $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W_p\big)$, we obtain

$$d\big(R_p(u_{k+1}), p\big) \leqslant \varrho(p)\,\|u_{k+1}\| \leqslant \varrho_0 \left(\eta_0 r' + \frac{1}{\rho_0}\sigma + \frac{1}{1-\theta}\kappa s'\right) < \lambda_0,$$

which means $z_{k+1} = R_p(u_{k+1}) \in W_p$. Thus, there exists an element $v_{k+1} \in \iota^*\mathbb{B}_q$ with $R_q(v_{k+1}) = R_p(u_{k+1})$, the sequences (u_k) and (v_k) are well-defined. The rest of proof is similar to the one of Proposition 5.2. □

Remark 5.6. If the set-valued part Φ is invariant, i.e. $\Phi(x) \equiv K$ for a fixed set $K \subset \mathbb{R}^n$, conclusions of both Propositions 5.2 and 5.5 are valid without conditions imposed on the maps $R_p^{-1}R_q - P_\chi^{b,a}$. Indeed, since all parallel transports $P_\chi^{a,b}$ are linear isometry, the validity of metric regularity for Ψ_q implies the one for $\Psi_q P_\chi^{a,b}$, and vice versa. However, it is not difficult to verify that $\Psi_q P_\chi^{a,b} = \Psi_p + G_{\chi,p,q}$, so Theorem 3.3 and Theorem 3.4 in Section 3.1 can be used.

5.2 Convergence of Newton-Type Algorithm

We describe some notations that will be necessary in the sequel. Firstly, for simplicity, we use a convention that retraction R_z is well-defined when z varies in the manifold $\mathcal{M}$. (The map sending $(z, v) \in T\mathcal{M}$ to $R_z(v)$ is called a *global retraction* on $\mathcal{M}$.) Given an open subset Ω of $\mathcal{M}$, and let (λ_R, ι_R) be a normal pair on Ω in the sense of Proposition 2.22. Additionally, let $L : [0, +\infty) \longrightarrow [0, +\infty)$ be an increasing function. Global retraction R is said to be of an *equi-Lipschitz class* around a point $p \in \Omega$ with respect to L, written as $R \in \mathrm{ELC}_L\,(p)$, provided that the next assertion is fulfilled. If $d_{\mathfrak{R}}(p,q) \leqslant \lambda_R(p)$ and $d_{\mathfrak{R}}(p,q') \leqslant \lambda_R(p)$, the map $\Sigma_{q',q} := R_q^{-1}R_{q'} - P_\chi^{b,a}$ is Lipschitz continuous on the ball $\iota_R(p)\mathbb{B}_{q'}$ with a modulus $L(\ell(\chi))$ for any geodesic $\chi : [a,b] \longrightarrow \Omega$ having $\chi(a) = q$, $\chi(b) = q'$. Example 5.7 illustrates a situation where the property $R \in \mathrm{ELC}_L\left(p\right)$ might be satisfied.

Secondly, let $f : \Omega \longrightarrow \mathbb{R}^n$ be a smooth map and $\mathscr{D}f$ be its covariant derivative. Let $L : [0,+\infty) \longrightarrow [0,+\infty)$ be an increasing function. By the notation $\mathscr{D}f \in \mathscr{L}ip_L(\Omega)$ we mean for each geodesic path $\chi : [a,b] \longrightarrow \Omega$ it holds that $\left\|\mathscr{D}f(\chi(a)) - \mathscr{D}f(\chi(b))P_\chi^{a,b}\right\| \leqslant L\big(\ell(\chi)\big)$. When L is linear (i.e. $L = \kappa\,\mathrm{id}_{\mathbb{R}}$), such a property reduces the notion of Lipschitz continuity for covariant derivative used, e.g. in [34].

Example 5.7. Consider again the unit sphere $\mathbb{S}^{m-1}$ along with the retraction $R(p,u) = \frac{1}{\|p+u\|}(p+u)$ in Example 5.1. Define

$$L(r) = \frac{1-\cos r + \tan r}{\cos r - \tan r} + \frac{\sin r\,(1+\tan r)}{(\cos r - \tan r)^2} + \frac{1-\cos r+\sin r}{\cos r}, 0 \leqslant r < r_{\max} \tag{5.13}$$

where $r_{\max} \in \left(0, \frac{\pi}{2}\right)$ is solution of the equation $\cos r - \tan r = 0$. The numeric value is about $r_{\max} = \arcsin\left(\frac{\sqrt{5}-1}{2}\right) \approx 0.6662\ldots$

We will show that for each geodesic χ on $\mathbb{S}^{m-1}$ satisfying $\ell(\chi_{|[0,1]}) < r_{\max}$, the map $\Sigma_\chi := R_{\chi(0)}^{-1}R_{\chi(1)} - P_\chi^{1,0}$ is Lipschitz continuous on the unit ball of $T_{\chi(1)}\mathbb{S}^{m-1}$ together with a modulus $L\big(\ell(\chi)\big)$.
Indeed, according to [1], the geodesic χ can be written as follows

$$\chi(t) = \cos(\ell t)\,\chi(0) + \frac{\sin(\ell t)}{\ell}\dot\chi(0), \ell = \|\dot\chi(0)\| = \ell(\chi). \tag{5.14}$$

It is possible to check that for each t_0 and $v \in T_{\chi(t_0)}\mathbb{S}^{m-1}$, the field $\xi(v,t_0;t) = \left(\ell^{-2}\dot\chi(t_0)^Tv\right)\dot\chi(t) + \left[v - \left(\ell^{-2}\dot\chi(t_0)^Tv\right)\dot\chi(t_0)\right]$ adopts the following properties $\xi(v,t_0;t_0) = v$, $\xi(v,t_0;t) \in T_{\chi(t)}\mathbb{S}^{m-1}$ and $\nabla_{\dot\chi}\xi(v,t_0;\cdot) \equiv 0$. Thus, we obtain the following expression for parallel transport

$$P_\chi^{1,0}v = \left(\ell^{-2}\dot\chi(1)^Tv\right)\dot\chi(0) + \left[v - \left(\ell^{-2}\dot\chi(1)^Tv\right)\dot\chi(1)\right], v \in T_{\chi(1)}\mathbb{S}^{m-1}. \tag{5.15}$$

From the definition of R, one gets $R_p^{-1}(q) = \left(p^Tq\right)^{-1}q - p$ whenever $d_{\mathfrak{R}}(q,p) = \arccos\left(q^Tp\right) < \frac{\pi}{2}$. Consequently, by deriving the differentiation, we deduce

$$\begin{cases} d\left(R_p\right)_u(u') = \|p+u\|^{-1}u' - \|p+u\|^{-2}\left(u^Tu'\right)R_p(u); u,u' \in T_p\mathbb{S}^{m-1}, \\ d\left(R_p^{-1}\right)_q(v) = -\left(p^Tq\right)^{-2}\left(p^Tv\right)q + \left(p^Tq\right)^{-1}v; q^Tp > 0, v \in T_q\mathbb{S}^{m-1}. \end{cases} \tag{5.16}$$

Let $\Sigma_\chi = R^{-1}_{\chi(0)} R_{\chi(1)}$ and $w, w' \in T_{\chi(1)}\mathbb{S}^{m-1}$. Thanks to the chain rule, we have $D\Sigma_\chi(w) = d\left(R^{-1}_{\chi(0)}\right)_{R_{\chi(1)}(w)} d\left(R_{\chi(1)}\right)_w$. So, after including the second expression in (5.16), the differential $D\Sigma_\chi(w)$ satisfies

$$\begin{aligned}(D\Sigma_\chi(w))(w') = &-\left(\chi(0)^T R_{\chi(1)}(w)\right)^{-2}\left[\chi(0)^T\left(d\left(R_{\chi(1)}\right)_w(w')\right)\right] R_{\chi(1)}(w) \\ &+ \left(\chi(0)^T R_{\chi(1)}(w)\right)^{-1} d\left(R_{\chi(1)}\right)_w(w').\end{aligned}$$

Using again (5.16), we reach an explicit representation for $D\Sigma_\chi(w)$ as follows

$$\begin{aligned}(D\Sigma_\chi(w))(w') = &\left(\chi(0)^T R_{\chi(1)}(w)\right)^{-1}\|\chi(1)+w\|^{-1} w' \\ &- \left(\chi(0)^T R_{\chi(1)}(w)\right)^{-2}\|\chi(1)+w\|^{-1}\left(\chi(0)^T w'\right) R_{\chi(1)}(w).\end{aligned} \tag{5.17}$$

Observe that $R_{\chi(1)}(w) = \|\chi(1)+w\|^{-1}\left[(\cos\ell)\,\chi(0) + (\ell^{-1}\sin\ell)\,\dot\chi(0) + w\right]$, which yields $\chi(0)^T R_{\chi(1)}(w) = \|\chi(1)+w\|^{-1}\left(\cos\ell + \chi(0)^T w\right)$. Since $w \in T_{\chi(1)}\mathbb{S}^{m-1}$, it holds that $0 = \chi(1)^T w = \cos\ell\left(\chi(0)^T w\right) + \ell^{-1}\sin\ell\left(\dot\chi(0)^T w\right)$. Therefore,

$$\chi(0)^T w = -\frac{\sin\ell}{\ell\cos\ell}\dot\chi(0)^T w = -\frac{\tan\ell}{\ell}\dot\chi(0)^T w. \tag{5.18}$$

Similarly, we also have

$$\chi(0)^T w' = -\frac{\tan\ell}{\ell}\dot\chi(0)^T w'. \tag{5.19}$$

As a result,

$$(D\Sigma_\chi(w))(w') = (\cos\ell)\,\nu_1\chi(0) + \frac{\sin\ell}{\ell}\nu_1\dot\chi(0) + \nu_1 w + \nu_2 w', \tag{5.20}$$

where

$$\begin{cases}\nu_1 = \frac{\tan\ell}{\ell}\left[\cos\ell - \frac{\tan\ell}{\ell}\dot\chi(0)^T w\right]^{-2}\left(\dot\chi(0)^T w'\right), \\ \nu_2 = \left[\cos\ell - \frac{\tan\ell}{\ell}\dot\chi(0)^T w\right]^{-1}.\end{cases} \tag{5.21}$$

Taking into account (5.15), and using $\dot\chi(t) = -\ell\sin(\ell t)\,\chi(0) + \cos(\ell t)\,\dot\chi(0)$, we obtain

$$\begin{aligned}(D\mathcal{E}_\chi(w))(w') = (D\Sigma_\chi(w))(w') - P^{1,0}_\chi w' = &(\nu_2 - 1)w' \\ &+ (\nu_1\cos\ell - \nu_3)\,\chi(0) + \left(\frac{\sin\ell}{\ell}\nu_1 - \nu_4\right)\dot\chi(0) + \nu_1 w.\end{aligned} \tag{5.22}$$

Here, ν_3 and ν_4 are given by

$$\nu_3 = \ell^{-1}\sin\ell\,(\cos\ell + \sin\ell\tan\ell)\left(\dot\chi(0)^T w'\right), \tag{5.23a}$$

$$\nu_4 = \ell^{-2} \left(\cos\ell + \sin\ell\tan\ell\right)\left(1-\cos\ell\right)\left(\dot{\chi}(0)^T w'\right). \tag{5.23b}$$

Suppose $\ell < r_{\max}$ and $\|w\| < 1$. Then we infer $\nu_2^{-1} \geqslant \cos\ell - \frac{\tan\ell}{\ell}\|\dot{\chi}(0)\| = \cos\ell - \tan\ell > 0$ by (5.21). Invoking $\left|\dot{\chi}(0)^T w'\right| \leqslant \|\chi(0)\|\,\|w'\| = \ell\,\|w'\|$, the relations (5.21), (5.22), (5.23a) and (5.23b) give us

$$\|(D\varSigma_\chi(w))\,(w')\| \leqslant L(\ell)\,\|w'\|\,;\ \|w\| < 1, \ell(\chi) = \ell < r_{\max}. \tag{5.24}$$

In summary, the expected conclusion holds.

We are now in position to present the main results of this chapter. Local convergence analysis for algorithm (5.2) will be in the next theorem.

Theorem 5.8 (local analysis). *Let $p^* \in \mathcal{M}$ be a solution of problem* (5.1) *and let $L_1, L_2 : [0,+\infty) \longrightarrow [0,+\infty)$ be increasing continuous functions with $L_1(0) = L_2(0) = 0$. Suppose that $\mathscr{D}\varphi$ admits the relation $\mathscr{D}\varphi \in \mathscr{L}ip_{L_1}(\Omega^*)$ for some open neighborhood Ω^* of p^* while $R \in \mathrm{ELC}_{L_2}\left(p^*\right)$ in the sense described at the beginning of this section. We assume in addition that $\tau^* = \operatorname{Reg}\varPsi^*(0_{p^*}, -\varphi(p^*)) < +\infty$ for $\varPsi^*(\cdot) = \mathscr{D}\varphi(p^*)(\cdot) + (\varPhi R_{p^*})(\cdot)$. If $R \in \mathbf{URC}\big(\rho, \varrho, \delta, \Omega^*\big)$, $\inf_{z\in\Omega^*}\rho(z) > 0$, $\inf_{z\in\Omega^*}\delta(z) > 0$ and $\sup_{z\in\Omega^*}\varrho(z) < +\infty$, then there is a neighborhood U^* of p^* satisfying the following statement. Starting at $p_0 = p \in U^*$, algorithm* (5.2) *generates a sequence (p_k) converging to p^* at least superlinearly.*

The proof of this theorem needs some technical estimates. Proposition 5.9 is in this sense.

Proposition 5.9. *Keep in mind the assumptions of Theorem 5.8. Let $p \in \Omega^*$ and $u \in \iota_R(p^*)\mathbb{B}_p$ so that the retraction segment $\gamma(t) = R_p(tu)$, $t \in [0,1]$ lies inside the ball $\mathbb{B}_{\mathcal{M}}(p^*, \lambda_R(p^*))$. If the image $\gamma([0,1])$ is contained into a convex neighborhood of p^* then one has*

$$\begin{aligned}&\|\varphi(q) - [\varphi(p) + \mathscr{D}\varphi(p)(u)]\| \\ &\qquad \leqslant \|u\| \int_0^1 \Big\{L_1\big(a(t)\big)\left[L_2\big(a(t)\big)+1\right] + \mu_p L_2\big(a(t)\big)\Big\}\,dt,\end{aligned} \tag{5.25}$$

for $q = R_p(u)$, $a(t) = d\big(p,\gamma(t)\big)$ and $\mu_p = \|\mathscr{D}\varphi(p)\|$.

Proof. For shortness, let's abbreviate $\iota^* = \iota_R(p^*)$ and $\lambda^* = \lambda_R(p^*)$. Without loss of generality, we suppose $W^* = \mathbb{B}_{\mathcal{M}}(p^*, \lambda_R(p^*))$ is itself a convex neighborhood. That means, any two points $\gamma(t_1)$ and $\gamma(t_2)$ $(0 \leqslant t_1, t_2 \leqslant 1)$ can be joined to each other by a minimizing geodesic which totally lies in W^*. Considering the composition $f = \varphi \circ \gamma$, one has $f'(t) = \mathscr{D}\varphi(\gamma(t))\,(\gamma'(t))$. Thus,

$$\varphi(q) = f(1) = f(0) + \int_0^1 \mathscr{D}\varphi(\gamma(t))\,(\gamma'(t))\;dt = \varphi(p) + \int_0^1 \mathscr{D}\varphi(\gamma(t))\,(\gamma'(t))\;dt,$$

which implies that

$$\|\varphi(q) - [\varphi(p) + \mathscr{D}\varphi(p)(u)]\| \leqslant \int_0^1 \left\|\mathscr{D}\varphi(\gamma(t))\,(\gamma'(t)) - \mathscr{D}\varphi(p)(u)\right\| dt. \tag{5.26}$$

Let $\chi_t(\cdot)$ be a minimizing geodesic such that $\chi_t(0) = \gamma(t)$, $\chi_t(1) = p$ and $\chi_t([0,1]) \subset W^*$. We define $G_t = \mathscr{D}\varphi(\gamma(t))P^{1,0}_{\chi_t} - \mathscr{D}\varphi(p)$ and $\Sigma_t = R^{-1}_{\gamma(t)}R_p - P^{1,0}_{\chi_t}$. Then it holds that $L_2(\ell(\chi_t)) \in \operatorname{Lipmod}\big(\Sigma_t, \iota^*\mathbb{B}_p\big)$ whereas $\|G_t\| \leqslant L_1(\ell(\chi_t))$. Because of $R_p(tu) = \gamma(t)$ and $u \in \iota^*\mathbb{B}_p$, we get $\|(d\Sigma_t)\,(tu)\| \leqslant L_2(\ell(\chi_t))$. Through a simple computation

$$(d\Sigma_t)\,(tu) = \big(dR_{\gamma(t)}\big)^{-1}_{0_{\gamma(t)}}(dR_p)_{tu} - P^{1,0}_{\chi_t} = (dR_p)_{tu} - P^{1,0}_{\chi_t},$$

and this yields

$$\left\|\gamma'(t) - P^{1,0}_{\chi_t}(u)\right\| = \left\|(dR_p)_{tu}(u) - P^{1,0}_{\chi_t}(u)\right\| \leqslant L_2(\ell(\chi_t))\,\|u\|. \tag{5.27}$$

Nevertheless, it is possible to see that $(G_t + \mathscr{D}\varphi(p))\,P^{0,1}_{\chi_t} = \mathscr{D}\varphi(\gamma(t))$. Henceforth, taking into account (5.26), we deduce

$$\begin{aligned}\|\varphi(q) - [\varphi(p) + \mathscr{D}\varphi(p)(u)]\| \leqslant &\int_0^1 \left\|G_t(u)\right\| dt \\ &+ \int_0^1 \left\| \left[\big(G_t + \mathscr{D}\varphi(p)\big)P^{0,1}_{\chi_t}\right]\big(\gamma'(t) - P^{1,0}_{\chi_t}(u)\big) \right\| dt.\end{aligned} \tag{5.28}$$

Thanks to the minimality of χ_t, $\ell(\chi_t) = d\big(p, \gamma(t)\big)$. Thus, one obtains (5.25) by combining (5.27) with (5.28) . □

Proof of Theorem 5.8. At the begining, we denote $\rho^* = \inf_{z\in\Omega^*}\rho(z)$, $\varrho^* = \sup_{z\in\Omega^*}\varrho(z)$, $\eta^* = \varrho^*/\rho^* \geqslant 1$ to simplify the notations. Let $r > 0$ and $s > 0$ so that, Ψ^* is metrically regular on $\mathcal{V} = \mathbb{B}_{p^*} \times \mathbb{B}_{\mathbb{R}^n}(-\varphi(p^*), s)$ together with a modulus $\tau \geqslant \tau^*$. Next, we look for a positive number $\alpha^* \leqslant \lambda_R(p^*)$ such that $\mathbb{B}_{\mathcal{M}}(p^*, \alpha^*) \subset \Omega^*$ is a convex neighborhood at p^*. Put $\mu = \sup_{d(p,p^*)<\alpha^*}\|\mathscr{D}\varphi(p)\|$ and $L(t) = L_1(t)\,[L_2(t) + 1] + \mu L_2(t)$. The quantity μ is finite according to the fact $\mathscr{D}\varphi \in \mathscr{L}ip_{L_1}(\Omega^*)$. Making α^* smaller if necessary, we require the following constraints are fulfilled

$$2\varrho^*\tau L(\eta^*\alpha^*) < \rho^*, \tag{5.29a}$$
$$\frac{1}{\rho^*}\frac{\tau L(\eta^*\alpha^*)}{1-\tau\eta^* L(\eta^*\alpha^*)}\alpha^* < r, \tag{5.29b}$$
$$\frac{1}{\rho^*}\frac{L(\eta^*\alpha^*)}{1-\tau\eta^* L(\eta^*\alpha^*)}\alpha^* < s, \tag{5.29c}$$
$$\frac{1}{\rho^*}\left(1+\frac{\tau L(\eta^*\alpha^*)}{1-\tau\eta^* L(\eta^*\alpha^*)}\right)\alpha^* < \iota_R(p^*), \tag{5.29d}$$
$$\eta^*\frac{\tau L(\eta^*\alpha^*)}{1-\tau\eta^* L(\eta^*\alpha^*)}\alpha^* < \lambda_R(p^*). \tag{5.29e}$$

Let's pick a value

$$0 < \alpha \leqslant \min\left\{\frac{\rho^*}{\rho^*+\varrho^*}\alpha^*, \lambda_R(p^*), \iota_R(p^*)\right\} \tag{5.30}$$

and choose $U^* = \mathbb{B}_{\mathcal{M}}(p^*, \alpha)$.

Fixing $p_0 \neq p^*$ in U^* and we call by χ_0 the minimizing geodesic path joining $p^* = \chi_0(0)$ to $p_0 = \chi_0(1)$. Let $u_0^* \in T_{p_0}\mathcal{M}$ such that $\|u_0^*\| < \iota_R(p^*)$ and $R_{p_0}(u_0^*) = p^*$. We denote $\alpha_0 = d(p^*, p_0) < \alpha$, $x_0 = -\varphi(p^*) + \mathscr{D}\varphi(p^*)(u_0^*)$, $y_0 = -\varphi(p_0)$ and $\Psi_0(\cdot) = \mathscr{D}\varphi(p_0)(\cdot) + (\Phi\circ R_{p_0})(\cdot)$. Then it follows from the assumptions of Theorem 5.8 that

$$\|G_{\chi_0,p^*,p_0}\| \leqslant L_1(\alpha_0), L_2(\alpha_0) \in \operatorname{Lipmod}\big(\Sigma_{\chi_0,p_0,p^*}, \iota_R(p^*)\mathbb{B}_{p_0}\big),$$

in which the notations $G_{\chi_0,p^*,p_0} = \mathscr{D}\varphi(p^*) - \mathscr{D}\varphi(p_0)P_0$, $\Sigma_{\chi_0,p_0,p^*} = R_{p^*}^{-1}R_{p_0} - P_0^{-1}$, $P_0 = P_{\chi_0}^{0,1}$ were involved. Consider the retraction segment $\gamma_0(t) = R_{p_0}(tu_0)$ for $0 \leqslant t \leqslant 1$. Observing $\|u_0^*\| \leqslant \frac{1}{\rho^*}d\big(p_0, R_{p_0}(u_0^*)\big) = \frac{1}{\rho^*}\alpha_0$, so one has

$$a_0(t) := d\big(p_0, \gamma_0(t)\big) \leqslant \varrho^* t\,\|u_0^*\| \leqslant t\frac{\varrho^*}{\rho^*}\alpha_0 = \eta^*\alpha_0 t \leqslant \eta^*\alpha_0,\ \ 0 \leqslant t \leqslant 1.$$

Hence, $d\big(p^*, \gamma_0(t)\big) \leqslant (1+\eta^*)\alpha_0$. According to (5.30), the image $\gamma_0([0,1])$ thereby lies in $\mathbb{B}_{\mathcal{M}}(p^*, \alpha^*)$. By Proposition 5.9, we get

$$\begin{aligned}
\|y_0 - x_0\| &= \|\varphi\left(R_{p_0}(u_0^*)\right) - [\mathscr{D}\varphi(p^*)(u_0^*) + \varphi(p_0)]\| \\
&\leqslant \|u_0^*\| \int_0^1 \Big\{L_1\big(\eta^*\alpha_0 t\big)\left[L_2\big(\eta^*\alpha_0 t\big)+1\right] + \|\mathscr{D}\varphi(p^*)\|\, L_2\big(\eta^*\alpha_0 t\big)\Big\}\, dt \\
&\leqslant \frac{1}{\rho^*}\alpha_0 \frac{1}{\eta^*\alpha_0}\int_0^{\eta^*\alpha_0} \Big\{L_1(t)\left[L_2(t)+1\right] + \mu L_2(t)\Big\}\, dt
\end{aligned}$$

$$= \frac{1}{\varrho^*} \int_0^{\eta^* \alpha_0} L(t)\, dt.$$

In particular, it holds that

$$\|y_0 - x_0\| \leqslant \frac{1}{\varrho^*} \eta^* \alpha_0 L(\eta^* \alpha_0) = \frac{1}{\rho^*} \alpha_0 L(\eta^* \alpha_0). \tag{5.31}$$

Remind $L_1(\eta^*\alpha_0) + L_2(\eta^*\alpha_0)\mu^* \leqslant L(\eta^*\alpha_0)$, the following inequalities are fulfilled by the choice of α

$$\begin{cases} \theta_0 = \eta^* \tau [L_1(\eta^*\alpha_0) + L_2(\eta^*\alpha_0)\mu^*] & < 1, \\ \frac{\tau}{1-\theta_0} d_0 & < r, \\ \frac{1}{1-\theta_0} d_0 & < s, \\ \alpha_0/\rho^* + \frac{\tau}{1-\theta_0} d_0 & < \iota_R(p^*), \\ \varrho^* \frac{\tau}{1-\theta_0} d_0 & < \lambda_R(p^*), \end{cases}$$

for $d_0 := \frac{1}{\rho^*}\alpha_0 L(\eta^*\alpha_0)$ and $\mu^* = \|\mathscr{D}\varphi(p^*)\|$. Consequently, we are able to add some positive parameters $r_0 > 0$ and $s_0 > \frac{1}{\rho^*}\alpha_0 L(\eta^*\alpha_0)$ such that

$$\begin{aligned} \eta^* \frac{2}{1-\theta_0} r_0 + \frac{\tau}{1-\theta_0} s_0 &< r, \\ (\eta^*)^2 [L_1(\eta^*\alpha_0) + L_2(\eta^*\alpha_0)\mu^*] \frac{2}{1-\theta_0} r_0 + \frac{1}{1-\theta_0} s_0 &< s, \\ \alpha_0/\rho^* + \eta^* \frac{2}{1-\theta_0} r_0 + \frac{\tau}{1-\theta_0} s_0 &< \iota_R(p^*), \\ \varrho^* \left(\eta^* \frac{2}{1-\theta_0} r_0 + \frac{\tau}{1-\theta_0} s_0 \right) &< \lambda_R(p^*). \end{aligned}$$

Following the proof of Proposition 5.2, one has $\tau_0 \in \operatorname{Regmod}\big(\Psi_0, \mathcal{V}_0\big)$ for $\tau_0 = \frac{\tau}{1-\theta_0}$ and $\mathcal{V}_0 = \big\{(u,x) : \|u - u_0^*\| < r_0, \|x - x_0\| < s_0\big\}$. By virtue of (5.31), the pair (u_0^*, y_0) is in $\mathcal{V}_0$. Thus, after applying the metric regularity property to Ψ_0, we can find $u_0 \in \Psi_0^{-1}(y_0)$ such that

$$\begin{aligned} \|u_0^* - u_0\| &\leqslant \tau_0 d\big(y_0, \Psi_0(u_0^*)\big) \leqslant \tau_0 \|y_0 - x_0\| \leqslant \frac{\tau}{1-\theta_0} \frac{1}{\rho^*} \alpha_0 L(\eta^*\alpha_0) \\ &\leqslant \frac{\tau L(\eta^*\alpha_0)}{\rho^* - \varrho^* \tau L(\eta^*\alpha_0)} \alpha_0. \end{aligned} \tag{5.32}$$

To continue the construction, we define $p_1 = R_{p_0}(u_0)$. Based on the choice of u_0, the inclusion $-\varphi(p_0) = y_0 \in \Psi_0(u_0)$ is evident. But it can be rewritten in an equivalent form

$$0 \in \varphi(p_0) + \mathscr{D}\varphi(p_0)(u_0) + (\Phi \circ R_{p_0})(u_0).$$

That is, p_1 satisfies (5.2).

Thanks to the triangle inequality in $T_{p_0}\mathcal{M}$, it holds that $\|u_0\| \leqslant \|u_0^*\| + \|u_0 - u_0^*\|$. Recalling $R_{p_0}(u_0^*) = p^*$ and $u_0^* \in \iota^*\mathbb{B}_{p^*}$, it is possible to estimate the value of $\|u_0^*\|$ as follows $\|u_0^*\| \leqslant \frac{1}{\rho(p_0)} d(p_0, p^*) \leqslant \frac{1}{\rho^*}\alpha_0$. By taking into account (5.29d), (5.30) and (5.32), we arrive

$$\|u_0\| \leqslant \frac{1}{\rho^*}\alpha_0 + \frac{\tau L(\eta^*\alpha_0)}{\rho^* - \varrho^*\tau L(\eta^*\alpha_0)}\alpha_0 = \frac{1}{\rho^*}\left(1 + \frac{\tau L(\eta^*\alpha_0)}{1 - \eta^*\tau L(\eta^*\alpha_0)}\right)\alpha_0 < \iota_R(p^*).$$

Therefore,

$$d(p^*, p_1) \leqslant \varrho^* \|u_0^* - u_0\| \leqslant \frac{\varrho^*\tau L(\eta^*\alpha_0)}{\rho^* - \varrho^*\tau L(\eta^*\alpha_0)}\alpha_0. \tag{5.33}$$

Due to (5.33) and (5.29a), $d(p^*, p_1) \leqslant \alpha_0$, and the new point p_1 belongs to U^*. So, we can apply all arguments above into p_1 instead of p_0 and continue the current process. As a result, algorithm (5.2) produces a sequence (p_k) in U^* which satisfies the relation

$$d(p^*, p_{k+1}) \leqslant \frac{\varrho^*\tau L(\eta^*\alpha_k)}{\rho^* - \varrho^*\tau L(\eta^*\alpha_k)} d(p^*, p_k), k = 0, 1, \ldots \tag{5.34}$$

Here, α_k indicates the quantity $d(p^*, p_k)$. If $p_k = p^*$ at some index k, then from (5.34) we have $p_j = p^*$ for $j \geqslant k$. Otherwise, taking into account $L(t) \to 0$ as $t \to 0$, (5.34) shows that $\limsup\limits_{k\to\infty} \frac{d(p^*, p_{k+1})}{d(p^*, p_k)} = 0$. This completes our proof. □

In view of (5.34), the rate of convergence seems to concern the behavior of real function L around the origin. If a stronger hypothesis is simultaneously imposed on L_1 and L_2, then it might be possible to refine the conclusion of Theorem 5.8 slightly. The next statement is in this sense.

Theorem 5.10. *Involving all assumptions of Theorem 5.8. Suppose in addition that $L_1(t) \overset{t\to 0}{=} O(t)$ and $L_2(t) \overset{t\to 0}{=} O(t)$. Then the resulting sequence (p_k) obtained in proof of the preceding theorem converges quadratically to p^*. Here, notation $L_j(t) \overset{t\to 0}{=} O(t)$ means $\limsup\limits_{t\to 0} \frac{L_j(t)}{t} < +\infty$.*

Proof. The proof is totally analogous to the one of Theorem 5.8. Quadratic convergence is inferred from (5.34), just substitute $\limsup\limits_{t\to 0} L(t) = 0$ with the fact that $\limsup\limits_{t\to 0} \frac{L(t)}{t} < +\infty$. □

Remark 5.11. In [4], the authors have proved a similar result as Theorem 5.10 in the case where both L_1 and L_2 are C^1. It is clear that in such a situation the conditions $L_j(t) \overset{t\to 0}{=} O(t)$ are fulfilled. This permits us to subsume the corresponding statement attained in [4] as a particular case of the one in Theorem 5.10.

To end up the current section, we introduce a result of global type compared to Theorem 5.10.

Theorem 5.12 (semi-local analysis)**.** *Keep in mind φ, Φ, L_1 and L_2 as similar as in Theorem 5.8, where p and Ω are replaced by p^* and Ω^* respectively. Let $\Psi(\cdot) := \mathscr{D}\varphi(p)(\cdot) + (\Phi R_p)(\cdot)$ and suppose that $\tau \in \operatorname{Regmod}\big(\Psi, V\big)$, in which*

$$V = \big\{(u,x) : \|u\| \leqslant r, d\big(x, \Psi(u)\big) \leqslant s\big\}$$

and the parameters r, s, τ are positive. Being stronger than in Theorem 5.8, we restrict on the case where both L_1, L_2 are C^1 only. Put $K_j(r) = \sup_{0\leqslant t\leqslant r} \left|L_j'(t)\right|$, and $\mu = \sup_{q\in\Omega} \|\mathscr{D}\varphi(q)\|$. Let's define

$$K = \left(\frac{\rho^*}{\varrho^*}\right)^2 \tau\Big\{K_1(r)[1 + L_2(r)] + \mu K_2(r)\Big\}, \beta(\tau, p) = \tau\rho^* d\big(0, \varphi(p) + \Phi(p)\big),$$

in which $\rho^* = \inf_{z\in\Omega} \rho(z)$, $\varrho^* = \sup_{z\in\Omega} \varrho(z)$. We assume the conditions below are valid as well:

(i) $W := \mathbb{B}_{\mathcal{M}}(p, \lambda_R(p))$ is a convex neighborhood;

(ii) $d\big(0, \varphi(p) + \Phi(p)\big) < s$;

(iii) $\alpha := 2K\beta(\tau, p) \leqslant 1$;

(iv) $\nu\beta(\tau, p) < \min\left\{r, \frac{(\rho^*)^2}{\varrho^*}r, \frac{(\rho^*)^2}{\varrho^*}\delta^*, \left(\frac{\rho^*}{\varrho^*}\right)^2 \lambda_R(p), \frac{(\rho^*)^2}{\varrho^*}\iota_R(p)\right\}$ for $\nu := \frac{2}{1+\sqrt{1-\alpha}}$ and $\delta^* := \inf_{z\in\Omega} \delta(z)$.

Under those hypotheses, problem (5.1) admits a solution p^* such that $d(p, p^*) \leqslant \nu\beta(\tau, p)$. Alternatively, starting at $p_0 = p$, algorithm (5.2) produces a sequence $p_k \in \mathcal{M}$ fulfilling the estimation

$$\begin{cases} d(p_k, p^*) \leqslant \frac{4\sqrt{1-\alpha}}{\alpha}\frac{b^{2^k}}{1-b^{2^k}}\beta, & \text{if } \alpha < 1, \\ d(p_k, p^*) \leqslant 2^{-k+1}\beta, & \text{if } \alpha = 1, \end{cases} \tag{5.35}$$

with $b := \frac{1-\sqrt{1-\alpha}}{1+\sqrt{1-\alpha}}$.

Proof. We briefly write $\lambda = \lambda_R(p)$, $\iota = \iota_R(p)$ and $\eta^* = \varrho^*/\rho^*$. As similar as in Theorem 3.9, the case $\beta = \beta(\tau, p) = 0$ is trivial. We separate the proof in several parts.

• Majorizing equation.

The majorizing function is a quadratic polynomial $\omega(t) = \frac{1}{2}Kt^2 - t + \beta$. ω has two positive roots of which $t^* = \frac{2}{1+\sqrt{1-\alpha}}\beta$ is the smallest one. We also apply the result in [40] to conclude that, under initial datum $t_0 = 0$, the classical Newton method generates a sequence $t_{k+1} = t_k - \omega'(t_k)^{-1}\omega(t_k)$ being strictly increasing. Furthermore, if $\alpha < 1$ it satisfies the error bound

$$\begin{cases} t^* - t_k \leqslant \frac{4\sqrt{1-\alpha}}{\alpha}\frac{b^{2^k}}{1-b^{2^k}}(t_1 - t_0) = \frac{4\sqrt{1-\alpha}}{\alpha}\frac{b^{2^k}}{1-b^{2^k}}\beta, \\ \frac{2(t_{k+1}-t_k)}{1+\sqrt{1+4b^{2^k}\left(1+b^{2^k}\right)^{-2}}} \leqslant t^* - t_k \leqslant b^{2^{k-1}}(t_k - t_{k-1}). \end{cases} \tag{5.36}$$

When $\alpha = 1$, (5.36) is replaced by

$$\begin{cases} t^* - t_k \leqslant 2^{-k+1}(t_1 - t_0) = 2^{-k+1}\beta, \\ 2\left(\sqrt{2}-1\right)(t_{k+1} - t_k) \leqslant t^* - t_k \leqslant t_k - t_{k-1}. \end{cases} \tag{5.37}$$

In particular, by induction with respect to k, we can see that $t_{k+1} - t_k \leqslant \beta$.

• Construction of the approximating solution.

Let us start with a guess point $p_0 = p$ satisfying all conditions in statement of Theorem 5.12. We shall look for $p_1, p_2, \ldots$ such that

$$d(p_i, p_{i+1}) \leqslant \eta^*\beta_i, \text{ for } \beta_i := t_{i+1} - t_i. \tag{5.38}$$

It is sufficient to carry out the induction step only, because p_1 can be analogously obtained from the metric regularity of mapping $\Psi_0(\cdot) := \mathscr{D}\varphi(p_0)(\cdot) + (\Phi R_{p_0})(\cdot)$ as in the proof of Theorem 3.9. Assume that $p_1, \ldots, p_k \in W$ and $u_0 \in \iota\mathbb{B}_{p_0}, \ldots, u_{k-1} \in \iota\mathbb{B}_{p_{k-1}}$ are known. Moreover, the iterations $p_1, \ldots, p_k$ are supposed to fulfill both (5.2) and (5.38), while each tangent vectors u_i $(i = 0, \ldots, k-1)$ fulfills

$$\|u_i\| \leqslant (\rho^*)^{-1}\beta_i. \tag{5.39}$$

We want to seek p_{k+1} such that $d(p_k, p_{k+1}) \leqslant \eta^*(t_{k+1} - t_k)$. For this goal, let us denote $\Psi_k(\cdot) := \mathscr{D}\varphi(p_k)(\cdot) + (\Phi R_{p_k})(\cdot)$ and $x_k = -\varphi(p_k)$. Since p_k obeys the framework of (5.2), inclusion $0 \in \varphi(p_{k-1}) + \mathscr{D}\varphi(p_{k-1})(u_{k-1}) + \Phi(p_k)$ is straightforward. In other words, $-\varphi(p_{k-1}) - \mathscr{D}\varphi(p_{k-1})(u_{k-1})$ is an element of the set $\Phi(p_k)$. Therefore,

$$d\big(x_k, \Psi_k(0_{p_k})\big) = d\big(x_k, \Phi(p_k)\big) \leqslant \|\varphi(p_k) - [\varphi(p_{k-1}) + \mathscr{D}\varphi(p_{k-1})(u_{k-1})]\| . \quad (5.40)$$

We have known in the previous part $\beta_{k-1} \leqslant \beta$, so $\|u_{k-1}\| \leqslant \frac{1}{\rho^*}\beta$. Define $\gamma_{k-1}(t) := R_{p_{k-1}}(tu_{k-1})$ and $a_{k-1}(t) := d\big(p_{k-1}, \gamma_{k-1}(t)\big)$. The function a_{k-1} can be dominated in the interval $[0,1]$ as follows $a_{k-1}(t) \leqslant \varrho^* \|u_{k-1}\| \, t \leqslant \eta^* \beta_{k-1} t$. Consequently,

$$\begin{aligned} d\big(p, \gamma_{k-1}(t)\big) &\leqslant \sum_{i=0}^{k-2} d(p_i, p_{i+1}) + a_{k-1}(t)\big) \leqslant \eta^* \sum_{i=0}^{k-2} (t_{i+1} - t_i) + \eta^* \beta_{k-1} t \\ &\leqslant \eta^*(t_{k-1} + \beta_{k-1}) = \eta^* t_k < \eta^* t^* < \lambda, \end{aligned}$$

which means that γ_{k-1} is not out of the neighborhood W. Then, repeating the Proposition 5.9 and letting $p = p_{k-1}$, we deduce

$$\begin{aligned} &\|\varphi(p_k) - [\varphi(p_{k-1}) + \mathscr{D}\varphi(p_{k-1})(u_{k-1})]\| \\ &\leqslant \|u_{k-1}\| \int_0^1 \Big\{ L_1\big(a_{k-1}(t)\big) \big[L_2\big(a_{k-1}(t)\big) + 1\big] + \|\mathscr{D}\varphi(p_{k-1})\| \, L_2\big(a_{k-1}(t)\big) \Big\} \, dt \\ &\leqslant \frac{1}{\rho^*} \beta_{k-1} \int_0^1 \Big\{ L_1\big(\eta^* \beta_{k-1} t\big) \big[L_2\big(\eta^* \beta_{k-1} t\big) + 1\big] + \mu L_2\big(\eta^* \beta_{k-1} t\big) \Big\} \, dt \\ &\leqslant \frac{1}{\varrho^*} \int_0^{\eta^* \beta_{k-1}} \Big\{ L_1(t) \, [L_2(t) + 1] + \mu L_2(t) \Big\} \, dt. \end{aligned}$$

Taking into account $\eta^* \beta_{k-1} \leqslant \eta^* \beta \leqslant r$, the fact that $0 \leqslant t \leqslant \eta^* \beta_{k-1}$ implies $L_j(t) \leqslant K_j(r) t$. As a result, we obtain

$$\begin{aligned} &\|\varphi(p_k) - [\varphi(p_{k-1}) + \mathscr{D}\varphi(p_{k-1})(u_{k-1})]\| \\ &\quad \leqslant \frac{1}{\varrho^*} \int_0^{\eta^* \beta_{k-1}} \Big\{ K_1(r) \, [L_2(r) + 1] + \mu K_2(r) \Big\} t \, dt = \frac{1}{2\tau \varrho^*} K \beta_{k-1}^2. \end{aligned} \quad (5.41)$$

In order to use Proposition 5.5, we first establish the following inequalities

$$\theta_k = \eta^* \tau \left[L_1\big(\eta^* t_k\big) + \mu_p L_2\big(\eta^* t_k\big) \right] < 1, \quad (5.42a)$$

$$\eta^* t_k / \rho^* + \frac{\tau}{1 - \theta_k} d_k < \min\left\{ \iota, \delta^*, r \right\}, \quad (5.42b)$$

$$\varrho^* \eta^* t_k / \rho^* + \varrho^* \frac{\tau}{1 - \theta_k} d_k < \lambda, \quad (5.42c)$$

for $d_k := \frac{1}{2\tau\varrho^*} K \beta_{k-1}^2$ and $\mu_p = \|\mathscr{D}\varphi(p)\| \leqslant \mu$. Indeed, based on the assumptions of Theorem 5.12, it is possible to check that the quantity $L_1\big(\eta^* t_k\big) + \mu_p L_2\big(\eta^* t_k\big)$

satisfies the estimations below

$$L_1\big(\eta^* t_k\big) + \mu_p L_2\big(\eta^* t_k\big) \leqslant \eta^* t_k[K_1(r) + \mu K_2(r)] \leqslant (\eta^*)^{-1} t_k \tau^{-1} K. \tag{5.43}$$

As a result, (5.42a) follows immediately from (5.43), since

$$(\eta^*)^{-1} t_k \tau^{-1} K < \tau^{-1}(\eta^*)^{-1} K t^* = \tau^{-1}(\eta^*)^{-1} \frac{\alpha}{1+\sqrt{1-\alpha}} \leqslant \tau^{-1}(\eta^*)^{-1}.$$

For (5.42b), we observe $K\beta_{k-1}^2 = \omega''(t_{k-1})\beta_{k-1}^2$. Expanding the polynomial ω at center t_{k-1}, and including relation $\omega(t_{k-1}) + \omega'(t_{k-1})\beta_{k-1} = 0$, we arrive

$$\begin{aligned} d_k &= \frac{1}{\tau\varrho^*}\left[\frac{1}{2}\omega''(t_{k-1})\beta_{k-1}^2\right] = \frac{1}{\tau\varrho^*}\Big\{\omega(t_k) - \big[\omega(t_{k-1}) + \omega'(t_{k-1})\beta_{k-1}\big]\Big\} \\ &= \frac{1}{\tau\varrho^*}\omega(t_k) \leqslant \frac{\eta^*}{\tau\rho^*}\omega(t_k). \end{aligned}$$

Due to (5.43), we conclude $\theta_k \leqslant K t_k = \omega'(t_k) + 1$. So, the left-hand side of (5.42b) does not exceed in $\eta^* t_k/\rho^* + \big(-\omega'(t_k)^{-1}\big)\frac{\eta^*}{\rho^*}\omega(t_k)$, which gives us

$$\eta^* t_k/\rho^* + \frac{\tau}{1-\theta_k} d_k \leqslant \frac{\varrho^*}{(\rho^*)^2}(t_k + \beta_k) = \frac{\varrho^*}{(\rho^*)^2} t_{k+1} < \frac{\varrho^*}{(\rho^*)^2} t^* \leqslant \min\big\{\iota, \delta^*, r\big\}.$$

Similarly, (5.42c) is verified as follows

$$\begin{aligned} \varrho^*\eta^* t_k/\rho^* + \varrho^* \frac{\tau}{1-\theta_k} d_k &\leqslant (\eta^*)^2 \left(t_k - \omega'(t_k)^{-1}\omega(t_k)\right) = (\eta^*)^2 t_{k+1} \\ &< (\eta^*)^2 t^* = (\eta^*)^2 \frac{2}{1+\sqrt{1-\alpha}}\beta \\ &= (\eta^*)^2 \nu\beta < \lambda. \end{aligned}$$

In summary, we can select some parameters $r_k > 0$ and $s_k > d_k$ such that

$$\begin{cases} \eta^* t_k/\rho^* + \varrho^* r_k/\rho^* + \frac{\tau}{1-\theta_k} s_k < \min\big\{\iota, \delta^*, r\big\}, \\ (\varrho^*)^2 r_k/\rho^* + \varrho^*\eta^* t_k/\rho^* + \varrho^* \frac{\tau}{1-\theta_k} s_k < \lambda. \end{cases}$$

Because of $d(p, p_k) \leqslant \sum_{i=0}^{k-1} d(p_i, p_{i+1}) \leqslant \eta^* \sum_{i=0}^{k-1} (t_{i+1} - t_i) = \eta^* t_k$, Proposition 5.5 shows that, the mapping Ψ_k is metrically regular with a modulus $\tau_k = \eta^* \frac{1}{1-\theta_k}$ on the set

$$V_k = \Big\{(u, x) : \|u\| \leqslant r_k, d\big(x, \Psi_k(u)\big) \leqslant s_k\Big\}.$$

According to (5.40) and (5.41), the pair $(0_{p_k}, x_k)$ belongs to V_k. Consequently, we obtain $d\left(0_{p_k}, \Psi_k^{-1}(x_k)\right) \leqslant \tau_k d\big(x_k, \Psi_k(0_{p_k})\big) \leqslant \tau_k d_k$. Recalling the previous evaluations $d_k = \frac{1}{\tau \varrho^*}\omega(t_k)$ and $\theta_k \leqslant K t_k = \omega'(t_k) + 1$, the closed set $\Psi_k^{-1}(x_k)$ contains one element, written by u_k, which satisfies

$$\|u_k\| = d\left(0_{p_k}, \Psi_k^{-1}(x_k)\right) \leqslant \tau_k d_k \leqslant \frac{1}{\rho^*}\left[-\omega'(t_k)^{-1}\omega(t_k)\right] = \frac{1}{\rho^*}(t_{k+1} - t_k).$$

Let's define $p_{k+1} = R_{p_k}(u_k)$. Since $\beta_k \leqslant \beta$, we derive from the assumption (iv) that

$$\|u_k\| \leqslant \frac{1}{\rho^*}\beta \leqslant \frac{1}{\rho^*}\frac{1+\sqrt{1-\alpha}}{2}\frac{(\rho^*)^2}{\varrho^*}\iota_R(p) < \iota_R(p).$$

Hence, the fact $R \in \mathbf{URC}\big(\rho, \varrho, \delta, W\big)$ could be applicable. Specifically, it yields

$$d(p_k, p_{k+1}) = d\big(p_k, R_{p_k}(u_k)\big) \leqslant \varrho^* \|u_k\| \leqslant \eta^* \beta_k,$$

so p_{k+1} obeys (5.38). Due to the triangle inequality, we find

$$d(p, p_{k+1}) \leqslant \sum_{i=0}^{k} d(p_i, p_{i+1}) \leqslant \eta^* \sum_{i=0}^{k} (t_{i+1} - t_i) = \eta^* t_{k+1} < \eta^* t^* \leqslant r.$$

Ultimately, p_{k+1} is still in the neighborhood W. The induction step is thereby completed.

- Convergence and error bounds.

Using relation (5.38) many times, we get

$$d(p_k, p_{k+j}) \leqslant \sum_{i=0}^{j-1} d(p_{k+i}, p_{k+i+1}) \leqslant \eta^* \sum_{i=0}^{j-1} (t_{k+i+1} - t_{k+i}) = \eta^*\left(t_{k+j} - t_k\right). \quad (5.44)$$

Based on (5.44), we conclude that (p_k) is a Cauchy sequence. Thus, there exists the limit $p^* = \lim_{k\to\infty} p_k$ in $\mathcal{M}$. By letting $j \to \infty$ in (5.44), and invoking both error bounds (5.36) and (5.37), we obtain (5.35).

At last, we claim $0 \in \varphi(p^*) + \Phi(p^*)$. In fact, recall that

$$0 \in \varphi(p_k) + \mathscr{D}\varphi(p_k)(u_k) + \left(\Phi R_{p_k}\right)(u_k) \quad (5.45)$$

holds for every index k. According to the convexity, there is a minimizing geodesic segment χ_k whose image belongs to W with $\chi_k(0) = p^*$ and $\chi_k(1) = p_k$. Let $G_k = \mathscr{D}\varphi(p^*) - \mathscr{D}\varphi(p_k)P_{\chi_k}^{0,1}$ be a linear operator. Thanks to the hypotheses of Theorem 5.12, the norm of G_k does not exceed in $L_1(\ell(\chi_k)) = L_1\big(d(p^*, p_k)\big)$. Since $\mathscr{D}\varphi(p_k) = \left(-G_k + \mathscr{D}\varphi(p^*)\right)P_{\chi_k}^{1,0}$, the sequence of linear operators $\{\mathscr{D}\varphi(p_k)\}$

has bounded norms. Thus,

$$\|\mathscr{D}\varphi(p_k)(u_k)\| \leqslant \|\mathscr{D}\varphi(p_k)\| \, \|u_k\| \leqslant \frac{1}{\rho^*} \|\mathscr{D}\varphi(p_k)\| \, (t_{k+1} - t_k) \xrightarrow{k\to\infty} 0.$$

Let us pass to the limit as $k \to \infty$ in (5.45), we deduce $0 \in \varphi(p^*) + \Phi(p^*)$. The proof of Theorem 5.12 is thereby completed. □

Remark 5.13. In the three Theorems 5.8, 5.10 and 5.12, the existence and the convergence of a Newton sequence (p_k) depend upon the assumptions related to covariant derivative $\mathscr{D}f$ as well as the retraction R. When the higher order covariant derivatives $\mathscr{D}^k f$ (see, e.g. [16, 82]) are included in the hypotheses of those theorems, we can obtain $\mathscr{D}\varphi \in \mathscr{L}ip_{L_1}(\Omega^*)$ from the informations on $\mathscr{D}^k f$. Therefore, it is possible to achieve some new versions of both Theorems 5.10 and 5.12 under conditions of type Kantorovich and Smale.

In the case $F(x) \equiv C$ and $R = \mathrm{Exp}$, Theorem 5.10 and Theorem 5.12 can be slightly improved. Indeed, as noticed in Remark 5.6, it is sufficient to adapt only the constraint of $\mathscr{D}f$ in order to obtain the conclusion of both Propositions 5.2 and 5.5. On the other hand, by letting $L = L_1$, the essential estimation in Proposition 5.9 still holds, since $P_\chi^{0,t}(\chi'(0)) = \chi'(t)$ for any arbitrary geodesic χ. Consequently, Kantorovich-type versions of Theorem 5.10 and Theorem 5.12 like in [82] can be recovered.

Now, we keep assuming $\Phi(p) \equiv C$ in Theorem 5.12. Furthermore, by interchanging the transportations $\mathcal{T}_R^{p,q}$ with parallelism $P_\chi^{a,b}$ and considering φ along the retraction curve $\gamma(t) = R_p(tu)$, (5.25) in Proposition 5.9 simply reads

$$\|\varphi(q) - [\varphi(p) + \mathscr{D}\varphi(p)(u)]\| \leqslant \|u\| \int_0^1 L_1\big(d\big(p, \gamma(t)\big)\big)\, dt. \tag{5.46}$$

Let us omit all conditions related to L_2, and involve just the ones for $L = L_1$ in Theorem 5.12. If the family of transportations $\mathcal{T}_R^{p,q}$ are assumed to be invertible and bounded norm, then we can recover Kantorovich-type results for retraction Newton's algorithm (5.2). Those might be viewed as extensions of the corresponding ones studied in the work [82].

5.3 An Example of Numerical Application

We illustrate the applicability of the preceding results by considering the problem of solving numerically a simple inclusion defined on 1-sphere $\mathbb{S}^1$ (see more details about unit sphere in Example 5.1). Choose $\varphi(p) = p_1 + p_2^2$ and $\Phi(p) = \{-p_1^2, p_1\}$. Furthermore, we also use the retraction $R_p(u) = \|p+u\|^{-1}(p+u)$ as in both Examples 5.1 and 5.7. Figure 5.3 describes geometrically such a retraction.

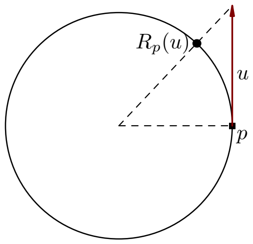


Fig. 5.3: Illustration of the sphere $\mathbb{S}^1$ and retraction R

In this situation, it is possible to set $\lambda_R \equiv \frac{\pi}{8}$ and $\iota_R \equiv 1$. Following Example 5.1, one has $R \in \mathbf{URC}(\rho, \varrho, \delta, \mathbb{S}^1)$, where $\rho(\cdot) \equiv 0.5$, $\varrho(\cdot) \equiv \pi/2$ and $\delta(\cdot) = \iota_R(\cdot)$. In addition, according to Example 5.7, one can check that $R \in \mathrm{ELC}_{L_2}(p)$ for any point $p \in \mathbb{S}^1$, in which $L_2 : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is the following function

$$L_2(r) = \begin{cases} \frac{1-\cos r+\tan r}{\cos r - \tan r} + \frac{\sin r(1+\tan r)}{(\cos r-\tan r)^2} + \frac{1-\cos r+\sin r}{\cos r}, & \text{if } r < r_{\max} \\ +\infty, & \text{otherwise,} \end{cases} \tag{5.47}$$

and $r_{\max} = \arcsin\left(\frac{\sqrt{5}-1}{2}\right) \approx 0.6662\ldots$ Finally, with the choice of function

$$L_1(r) = (3 + 2\cos r + \sin r)(\sin r + 1 - \cos r), \tag{5.48}$$

the hypothesis $\mathscr{D}\varphi \in \mathscr{L}ip_{L_1}(\mathbb{S}^1)$ is valid as well. Indeed, let $\gamma(t) = \cos(\ell t)\gamma(0) + \ell^{-1}\sin(\ell t)\dot\gamma(0)$ be an arbitrary geodesic on $\mathbb{S}^1$. Based on the arguments in Example 5.7, parallel transport along γ admits an explicit expression below

$$\begin{aligned} P_\gamma^{0,1} v &= \left(\ell^{-2}\dot\gamma(0)^T v\right)\dot\gamma(1) + \left[v - \left(\ell^{-2}\dot\gamma(0)^T v\right)\dot\gamma(0)\right] \\ &= v + \left(\dot\gamma(0)^T v\right)\left[\frac{\cos\ell - 1}{\ell^2}\dot\gamma(0) - \frac{\sin\ell}{\ell}\gamma(0)\right]. \end{aligned} \tag{5.49}$$

Consider the linear map $G_\gamma := \mathscr{D}\varphi(\gamma(0)) - \mathscr{D}\varphi(\gamma(1))P_\gamma^{0,1}$. We have $\mathscr{D}\varphi(p)(u) = u_1 + 2p_2u_2$ for $p = (p_1, p_2)^T \in \mathbb{S}^1$ and $u = (u_1, u_2)^T \in T_p\mathbb{S}^1$. Thus, for $v \in T_{\gamma(0)}\mathbb{S}^1$

$$\begin{aligned} G_\gamma(v) = v_1 + 2\gamma_2(0)v_2 &- \left\{v_1 + \left(\dot\gamma(0)^T v\right)\left[\frac{\cos\ell - 1}{\ell^2}\dot\gamma_1(0) - \frac{\sin\ell}{\ell}\gamma_1(0)\right]\right\} \\ &- 2\gamma_2(1)\left\{v_2 + \left(\dot\gamma(0)^T v\right)\left[\frac{\cos\ell - 1}{\ell^2}\dot\gamma_2(0) - \frac{\sin\ell}{\ell}\gamma_2(0)\right]\right\}, \end{aligned}$$

in which γ_1 and γ_2 indicate the components of γ. Taking into account $\gamma(1) = (\cos\ell)\,\gamma(0) + (\ell^{-1}\sin\ell)\,\dot\gamma(0)$, we deduce

$$G_\gamma(v) = G_1(v) + G_2(v) + G_3(v). \tag{5.50}$$

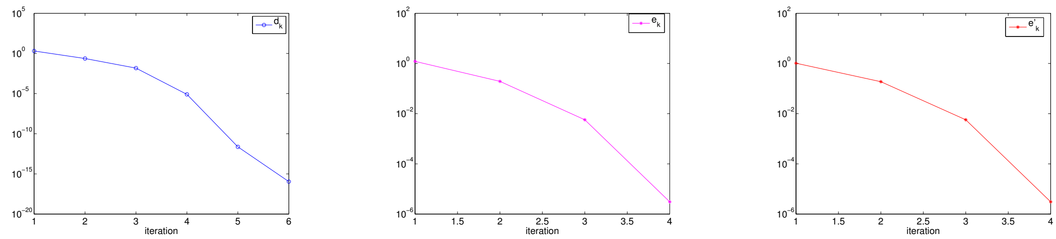
Here, the linear maps G_1, G_2 and G_3 are given by

$$G_1(v) = 2\left[(1-\cos\ell)\gamma_2(0) - \left(\ell^{-1}\sin\ell\right)\dot\gamma_2(0)\right] v_2, \tag{5.51}$$

$$G_2(v) = \frac{1-\cos\ell}{\ell^2}\left\{\dot\gamma_1(0) + 2\Big[(\cos\ell)\,\gamma_2(0) + \frac{\sin\ell}{\ell}\dot\gamma_2(0)\Big]\dot\gamma_2(0)\right\}\left(\dot\gamma(0)^T v\right), \tag{5.52}$$

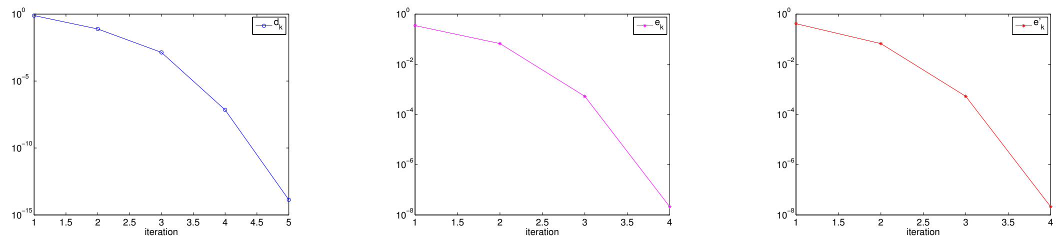
$$G_3(v) = \frac{\sin\ell}{\ell}\left\{\gamma_1(0) + 2\Big[(\cos\ell)\,\gamma_2(0) + \frac{\sin\ell}{\ell}\dot\gamma_2(0)\Big]\gamma_2(0)\right\}\left(\dot\gamma(0)^T v\right). \tag{5.53}$$

Noticing that $[\gamma_1(0)]^2 + [\gamma_2(0)]^2 = 1$ and $\ell = \|\dot\gamma(0)\| = [\dot\gamma_1(0)]^2 + [\dot\gamma_2(0)]^2$. According to (5.50), (5.51), (5.52) and (5.53), we obtain $\|G_\gamma\| \leqslant L_1(\ell)$, and this give us $\mathscr{D}\varphi \in \mathscr{L}ip_{L_1}(\mathbb{S}^1)$. Figures 5.4 and 5.5 sketch some practical computations corresponding to various guess points.



(a): $d_k = d\big(0, \varphi(p_k) + \Phi(p_k)\big)$ (b): $e_k = d\,(p_k, p_{k+1})$ (c): $e'_k = d\,(p_k, p_{kmax})$

Fig. 5.4: Numerical test: starting point $p_0 = (-0.9997, 0.0229)^T$



(a): $d_k = d\big(0, \varphi(p_k) + \Phi(p_k)\big)$ (b): $e_k = d\,(p_k, p_{k+1})$ (c): $e'_k = d\,(p_k, p_{kmax})$

Fig. 5.5: Numerical test: starting point $p_0 = (-0.1071, 0.9942)^T$

Bibliography

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Recent Advances in Optimization and its Applications in Engineering: The 14th Belgian-French-German Conference on Optimization*, chapter Optimization On Manifolds: Methods and Applications, pages 125–144. Springer Berlin Heidelberg, 2010.

[3] R. L. Adler, J. Dedieu, J. Y. Margulies, M. Martens, and M. Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA Journal of Numerical Analysis*, 22(3):359–390, 2002.

[4] S. Adly, H. V. Ngai, and V. V. Nguyen. Newton-type method for solving generalized equations on Riemannian manifolds. to appear.

[5] S. Adly, R. Cibulka, and H. V. Ngai. Newton's method for solving inclusions using set-valued approximations. *SIAM Journal on Optimization*, 25(1):159–184, 2015.

[6] S. Adly, H. V. Ngai, and V. V. Nguyen. Newton's method for solving generalized equations: Kantorovich's and Smale's approaches. *Journal of Mathematical Analysis and Applications*, 439(1):396–418, 2016.

[7] S. Adly, H. V. Ngai, and V. V. Nguyen. Stability of metric regularity with set-valued perturbations and application to Newton's method for solving generalized equations. submitted, 2016.

[8] F. Alvarez, J. Bolte, and J. Munier. A unifying local convergence result for Newton's method in Riemannian manifolds. *Foundations of Computational Mathematics*, 8(2):197–226, 2008.

[9] F. J. Aragón Artacho, A. Belyakov, A. L. Dontchev, and M. López. Local convergence of quasi-Newton methods under metric regularity. *Computational Optimization and Applications*, 58(1):225–247, 2013.

[10] A. B. Bakushinskii. A regularization algorithm based on the Newton–Kantorovich method for solving variational inequalities. *SSR Computational Mathematics and Mathematical Physics*, 16:16–23, 1976.

[11] R. Bhatia. *Positive Definite Matrices*, volume 107 of *Princeton Series in Applied Mathematics*. Princeton University Press, 2007.

[12] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation*. Springer New York, 1998.

[13] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.

[14] P. G. Ciarlet and C. Mardare. On the Newton–Kantorovich theorem. *Analysis and Applications*, 10(3):249–269, 2012.

[15] J.-P. Dedieu. *Points Fixes, Zéros et la Méthode de Newton*, volume 54 of *Mathématiques & Applications*. Springer Berlin Heidelberg, 2006.

[16] J.-P. Dedieu, P. Priouret, and G. Malajovich. Newton's method on Riemannian manifolds: covariant alpha theory. *IMA Journal of Numerical Analysis*, 23(3):395–419, 2003.

[17] W. Deren and Z. Fengguang. The theory of Smale's point estimation and its applications. *Journal of Computational and Applied Mathematics*, 60(1–2): 253–269, 1995.

[18] P. Deuflhard. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, volume 35 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg, 2011.

[19] S. Dias and G. Smirnov. On the Newton method for set-valued maps. *Nonlinear Analysis. Theory, Methods & Applications*, 75(3):1219–1230, 2012.

[20] M. P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser, 1992.

[21] A. L. Dontchev. Local analysis of Newton-type method based on partial linearization. In *Letters in Applied Mathematics*, volume 32, pages 295–306. American Mathematical Monthly, 1996.

[22] A. L. Dontchev. Local convergence of the Newton method for generalized equations. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, 322(4):327–331, 1996.

[23] A. L. Dontchev. Uniform convergence of the Newton method for Aubin continuous maps. *Serdica Mathematical Journal*, 22(3):283–296, 1996.

[24] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings: A View from Variational Analysis.* Springer Monographs in Mathematics. Springer New York, 2009.

[25] A. L. Dontchev and R. T. Rockafellar. Convergence of inexact Newton methods for generalized equations. *Mathematical Programming*, 139(1):115–137, 2013.

[26] A. L. Dontchev, A. S. Lewis, and R. T. Rockafellar. The radius of metric regularity. *Transactions of the American Mathematical Society*, 355:493–517, 2003.

[27] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1999.

[28] M. Fabian, P. Habala, P. Hájek, V. M. Santalucía, J. Pelant, and V. Zizler. *Functional Analysis and Infinite-Dimensional Geometry.* CMS Books in Mathematics. Springer New York, 2001.

[29] M. Fabian, P. Habala, P. Hájek, V. Montesinos, and V. Zizler. *Banach Space Theory: The Basis for Linear and Nonlinear Analysis.* CMS Books in Mathematics. Springer New York, 2011.

[30] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I.* Springer Series in Operations Research and Financial Engineering. Springer New York, 2003.

[31] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume II.* Springer Series in Operations Research and Financial Engineering. Springer New York, 2003.

[32] O. P. Ferreira. Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle. *IMA Journal of Numerical Analysis*, 29(3):746–759, 2009.

[33] O. P. Ferreira and R. C. M. Silva. Local convergence of Newton's method under a majorant condition in Riemannian manifolds. *IMA Journal of Numerical Analysis*, 32(4):1696–1713, 2012.

[34] O. P. Ferreira and B. F. Svaiter. Kantorovich's theorem on Newton's method in Riemannian manifolds. *Journal of Complexity*, 18(1):304–329, 2002.

[35] O. P. Ferreira and B. F. Svaiter. Kantorovich's majorants principle for Newton's method. *Computational Optimization and Applications*, 42(2):213–229, 2009.

[36] O. P. Ferreira, M. L. N. Gonçalves, and P. R. Oliveira. Convergence of the Gauss–Newton method for convex composite optimization under a majorant condition. *SIAM Journal on Optimization*, 23(3):1757–1783, 2013.

[37] M. C. Ferris and T. S. Munson. Complementarity problems in GAMS and the PATH solver. *Journal of Economic Dynamics and Control*, 24(2):165–188, 2000.

[38] A. Galántai. The theory of Newton's method. *Journal of Computational and Applied Mathematics*, 124(1–2):25 – 44, 2000.

[39] M. Gaydu and M. H. Geoffroy. A Newton iteration for differentiable set-valued maps. *Journal of Applied Mathematical Analysis and Applications*, 399(1):213–224, 2013.

[40] W. B. Gragg and R. A. Tapia. Optimal error bounds for the Newton–Kantorovich theorem. *SIAM Journal on Numerical Analysis*, 11(1):10–13, 1974.

[41] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Pure and Applied Mathematics. A Series of Monographs and Textbooks. Academic Press, INC, 1978.

[42] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer Berlin Heidelberg, 2001.

[43] S. Hosseini and M. R. Pouryayevali. Nonsmooth optimization techniques on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 158 (2):328–342, 2013.

[44] A. D. Ioffe. Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):501–558, 2000.

[45] A. D. Ioffe. On perturbation stability of metric regularity. *Set-Valued Analysis*, 9(1):101–109, 2001.

[46] A. F. Izmailov and M. V. Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2014.

[47] A. F. Izmailov and M. V. Solodov. Newton-type methods: A broader view. *Journal of Optimization Theory and Applications*, 164(2):577–620, 2015.

[48] A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov. The Josephy–Newton method for semismooth generalized equations and semismooth SQP for optimization. *Set-Valued and Variational Analysis*, 21(1):17–45, 2013.

[49] C. H. Jeffrey Pang. Generalized differentiation with positively homogeneous maps: Applications in set-valued analysis and metric regularity. *Mathematics of Operations Research*, 36(3):377–397, 2011.

[50] N. H. Josephy. Newton's method for generalized equations. Technical summary report, Mathematics Research Center, University of Wisconsin, Madison, 1979.

[51] L. V. Kantorovich and G. P. Akilov. *Functional Analysis*. Pergamon Press, 2 edition, 1982. Translated by Howard L. Silcock.

[52] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics, 1995.

[53] D. Klatte and B. Kummer. *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications*, volume 60 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 2002.

[54] N. Kollerstrom. Thomas Simpson and 'Newton's method of approximation': An enduring myth. *The British Journal for the History of Science*, 25(3): 347–354, 1992. URL http://www.jstor.org/stable/4027257.

[55] S. Lang. *Fundamentals of Differential Geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer New York, 1999.

[56] J. M. Lee. *Riemannian Manifolds: An Introduction to Curvature*, volume 176 of *Graduate Texts in Mathematics*. Springer New York, 1997.

[57] J. M. Lee. *Manifolds and Differential Geometry*, volume 107 of *Graduate Studies in Mathematics*. American Mathematical Society, 2009.

[58] C. Li, B. S. Mordukhovich, J. Wang, and J.-C. Yao. Weak sharp minima on Riemannian manifolds. *SIAM Journal on Optimization*, 21(4):1523–1560, 2011.

[59] B. S. Mordukhovich. Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions. *Transactions of the American Mathematical Society*, 340(1):1–35, 1993.

[60] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*, volume 330 of *Grundlehren der mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics)*. Springer Berlin Heidelberg, 2006.

[61] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation II: Applications*, volume 331 of *Grundlehren der mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics)*. Springer Berlin Heidelberg, 2006.

[62] B. S. Mordukhovich and Y. Shao. Differential characterizations of covering, metric regularity, and Lipschitzian properties of multifunctions between Banach spaces. *Nonlinear Analysis, Theory, Methods & Applications*, 25(12): 1401–1424, 1995.

[63] H. V. Ngai and M. Théra. Error bounds in metric spaces and application to the perturbation stability of metric regularity. *SIAM Journal on Optimization*, 19(1):1–20, 2008.

[64] J. M. Ortega. The Newton–Kantorovich theorem. *American Mathematical Monthly*, 75:658–660, 1968.

[65] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*, volume 30 of *Classic in Applied Mathematics*. Society for Industrial and Applied Mathematics, 2000.

[66] L. Qi and J. Sun. A nonsmooth version of newton's method. *Mathematical Programming*, 58(1):353–367, 1993.

[67] M. H. Rashid, S. H. Yu, C. Li, and S. Y. Wu. Convergence analysis of the Gauss–Newton-type method for Lipschitz-like mappings. *Journal of Optimization Theory and Applications*, 158:216–233, 2013.

[68] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds and their application to shape space. *SIAM Journal on Optimization*, 22(2): 596–627, 2012.

[69] S. M. Robinson. Extension of Newton's method to nonlinear functions with values in a cone. *Numerische Mathematik*, 19(4):341–347, 1972.

[70] S. M. Robinson. *Point-to-Set Maps and Mathematical Programming*, chapter Generalized equations and their solutions, part I: Basic theory, pages 128–141. Springer Berlin Heidelberg, 1979.

[71] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5(1):43–62, 1980.

[72] S. M. Robinson. *Optimality and Stability in Mathematical Programming*, chapter Generalized equations and their solutions, part II: Applications to nonlinear programming, pages 200–221. Springer Berlin Heidelberg, 1982.

[73] S. M. Robinson. Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17(3):691–714, 1992.

[74] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics)*. Springer Berlin Heidelberg, 1998.

[75] T. Sakai. *Riemannian Geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, 1996.

[76] W. Shen and C. Li. Smale's α-theory for inexact Newton methods under the γ-condition. *Journal of Mathematical Analysis and Applications*, 369(1):29 – 42, 2010.

[77] S. Smale. *The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics: Proceedings of a Symposium Held in Honor of Gail S. Young at the University of Wyoming, August 8–10, 1985*, chapter Newton's Method Estimates from Data at One Point, pages 185–196. Springer New York, 1986.

[78] S. T. Smith. Optimization techniques on Riemannian manifolds. In *Hamiltonian and gradient flows, algorithms and control*, volume 3 of *Fields Institute Communications*, pages 113–136. American Mathematical Society, 1994.

[79] C. Udrişte. *Convex Functions and Optimization Methods on Riemannian Manifolds*, volume 297 of *Mathematics and Its Applications*. Springer Netherlands, 1994.

[80] L. U. Uko. Generalized equations and the generalized Newton method. *Mathematical Programming*, 73(3):251–268, 1996.

[81] M. Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2011.

[82] J.-H. Wang, S. Huang, and C. Li. Extended Newton's method for mappings on Riemannian manifolds with values in a cone. *Taiwanese Journal of Mathematics*, 13(2B):633–656, 2009.

[83] X. Wang. Convergence of Newton's method and inverse function theorem in Banach space. *Mathematics of Computation*, 68(225):169–186, 1999.

[84] X. Wang. Convergence of Newton's method and uniqueness of the solution of equations in Banach space. *IMA Journal of Numerical Analysis*, 20(1): 123–134, 2000.

[85] B. Xiao and P. T. Harker. A nonsmooth Newton method for variational inequalities, I: theory. *Mathematical Programming*, 65:151–194, 1994.

[86] B. Xiao and P. T. Harker. A nonsmooth Newton method for variational inequalities, II: numerical results. *Mathematical Programming*, 65:195–216, 1994.

[87] J.-C. Yakoubsohn. Finding zeros of analytic functions: α-theory for secant type methods. *Journal of Complexity*, 15(2):239 – 281, 1999.

[88] P. P. Zabreiko and P. P. Zlepko. A generalization of the Newton-Kantorovich method on an equation with nondifferentiable operators. *Ukrainian Mathematical Journal*, 34(3):299–303, 1982. doi: 10.1007/BF01682124.

Titre thèse français: Méthode de Newton Revisitée pour Les Equations Généralisées.

Résumé : Le but de cette thèse est d'étudier la méthode de Newton pour résoudre numériquement les inclusions variationnelles, appelées aussi dans la littérature les équations généralisées. Ces problèmes engendrent en général des opérateurs multivoques. La première partie est dédiée à l'extension des approches de Kantorovich et la théorie (α, γ) de Smale (connues pour les équations non-linéaires classiques) au cas des inclusions variationnelles dans les espaces de Banach. Ceci a été rendu possible grâce aux développements récents des outils de l'analyse variationnelle et non-lisse tels que la régularité métrique. La seconde partie est consacrée à l'étude de méthodes numériques de type-Newton pour les inclusions variationnelles en utilisant la différentiabilité généralisée d'applications multivoques où nous proposons de linéariser à la fois les parties univoques (lisses) et multivoques (non-lisses). Nous avons montré que, sous des hypothèses sur les données du problème ainsi que le choix du point de départ, la suite générée par la méthode de Newton converge au moins linéairement vers une solution du problème de départ. La convergence superlinéaire peut-être obtenue en imposant plus de conditions sur l'approximation multivaluée. La dernière partie de cette thèse est consacrée à l'étude des équations généralisées dans les variétés Riemaniennes à valeurs dans des espaces euclidiens. Grâce à la relation entre la structure géométrique des variétés et les applications de rétractions, nous montrons que le schéma de Newton converge localement superlinéairement vers une solution du problème. La convergence quadratique (locale et semi-locale) peut-être obtenue avec des hypothèses de régularités sur les données du problème.

Mots clés : analyse variationnelle, régularité métrique, équations généralisées, méthode de Josephy-Newton, différenciation multivoque, méthode de Newton sur les variétés Riemaniennes, convergence linéaire/superlinéaire/quadratique.

Titre thèse anglais: Newton-type Methods for Solving Inclusions.

Abstract: This thesis is devoted to present some results in the scope of Newton-type methods applied for inclusion involving set-valued mappings. In the first part, we follow the Kantorovich's and/or Smale's approaches to study the convergence of Josephy-Newton method for generalized equation (GE) in Banach spaces. Such results can be viewed as an extension of the classical Kantorovich's theorem as well as Smale's (α, γ)-theory which were stated for nonlinear equations. The second part develops an algorithm using set-valued differentiation in order to solve GE. We proved that, under some suitable conditions imposed on the input data and the choice of the starting point, the algorithm produces a sequence converging at least linearly to a solution of considering GE. Moreover, by imposing some stronger assumptions related to the approximation of set-valued part, the proposed method converges locally superlinearly. The last part deals with inclusions involving maps defined on Riemannian manifolds whose values belong to an Euclidean space. Using the relationship between the geometric structure of manifolds and the retraction maps, we show that, our scheme converges locally superlinearly to a solution of the initial problem. With some more regularity assumptions on the data involved in the problem, the quadratic convergence (local and semi-local) can be ensured.

Keywords: variational analysis, metric regularity, generalized equations, Josephy-Newton method, set-valued differentiation, Riemannian Newton method, linear/superlinear/quadratic convergence.