Lecture Notes on

Nonlinear Inversion and Tomography:

I. Borehole Seismic Tomography

Developed from a Series of Lectures by

James G. Berryman
University of California
Lawrence Livermore National Laboratory
Livermore, CA 94550

Originally Presented at

Earth Resources Laboratory
Massachusetts Institute of Technology

July 9–30, 1990

Revised and Expanded

October, 1991

Chapter 2

Feasibility Analysis for Traveltime Inversion

The idea of using feasibility constraints in nonlinear programming problems is well established [Fiacco and McCormick, 1990]. However, it has only recently been realized that physical principles such as Fermat's principle actually lead to rigorous feasibility constraints for nonlinear inversion problems [Berryman, 1991]. The main practical difference between the standard analysis in nonlinear programming and the new analysis in nonlinear inversion is that, whereas the functions involved in nonlinear programming are often continuous, differentiable, and relatively easy to compute explicitly, the functionals in nonlinear inversion (*e.g.*, the traveltime functional) need not be continuous or differentiable and, furthermore, are very often comparatively difficult to compute. Feasibility constraints for inversion problems are implicit, rather than explicit.

We present the rigorous analysis here in a general setting, because it is actually quite easy to understand once we have introduced the concepts of convex function and convex set. This analysis is important because it will help to characterize the solution set for the inversion problem, and it will help to clarify questions about local and global minima of the inverson problem.

2.1 Feasibility Constraints Defined

Equation (1.5) assumes that P_i is a Fermat (least-time) path and leads to the equalities summarized in the vector-matrix equation $\mathbf{Ms} = \mathbf{t}$. Now let us suppose instead that P_i is a trial ray path which may or may not be the least-time path. Fermat's principle allows us to write

$$\int_{P_i} s(\mathbf{x}) dl^{P_i} \geq t_i, \tag{2.1}$$

where now t_i is the measured traveltime for source-receiver pair i. When we discretize (2.1) for cell or block models and all ray paths i, the resulting set of m inequalities may be written as

$$\mathbf{Ms} \geq \mathbf{t}. \tag{2.2}$$

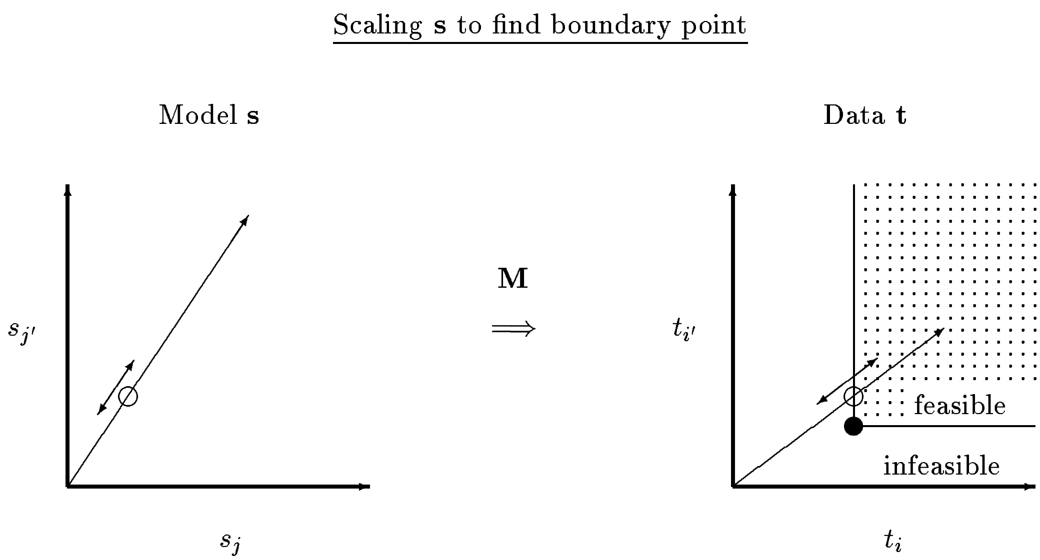

Figure 2.1: Feasible part of the model space is determined implicitly by the feasible part of the data space.

Equations (2.1) and (2.2) can be interpreted as a set of inequality constraints on the slowness model $\mathbf{s}$. When $\mathbf{s}$ obeys these m constraints, we say that $\mathbf{s}$ is *feasible*. When any of the constraints is violated, we say $\mathbf{s}$ is *infeasible*. The set of inequalities collectively will be called the *feasibility constraints*.

The concept of the feasibility constraint is quite straightforward in nonlinear programming problems [Fiacco and McCormick, 1990] whenever the constraints may be *explicitly* stated for the solution vector. However, in our inversion problems, an additional computation is required. Figure 2.1 shows that the feasibility constraints are *explicit* for the traveltime data vector, but they are only *implicit* (*i.e.*, they must be computed) for the slowness vector. This added degree of complication is unavoidable in the inversion problem, but nevertheless it is also very easily handled computationally with only very minor modifications of the usual nonlinear inversion algorithms.

2.2 Quick Review of Convexity

Here we define some mathematical concepts [Hardy, Littlewood, and Pólya, 1934] which will facilitate the discussion and analysis of feasible models. In the following, let $\mathcal{S}$ denote a linear vector space.

DEFINITION 2.2.1 (CONVEX SET) *A set $\mathcal{A} \subseteq \mathcal{S}$ is convex if, for every $\mathbf{s}_1, s_2 \in \mathcal{A}$ and every number $\lambda \in [0,1]$, we have $\lambda\mathbf{s}_1 + (1-\lambda)\mathbf{s}_2 \in \mathcal{A}$.*[1]

Examples of convex sets are

[1] $\mathcal{A} \subseteq \mathcal{S}$ means $\mathcal{A}$ is a *subset* of $\mathcal{S}$.

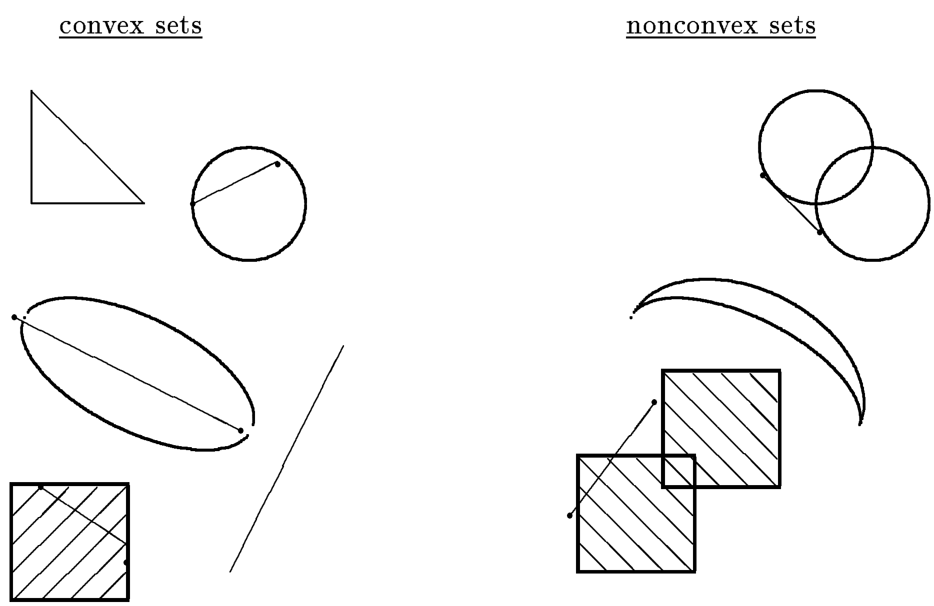

Figure 2.2: Examples of convex and nonconvex sets.

1. $\mathbf{R}$ (the real numbers).
2. $\mathbf{R}_+$ (the positive real numbers).
3. The positive n-tant $\mathbf{R}_+^n$; *i.e.*, the set of n-dimensional vectors whose components are all positive.
4. $C_+(\mathbf{R}^3)$ (the set of positive, continuous functions, $s(\mathbf{x}) > 0$, where $\mathbf{x} \in \mathbf{R}^3$).
5. A closed interval $[a, b]$ in $\mathbf{R}$.[2]
6. A hyperplane in $\mathbf{R}^n$; *i.e.*, vectors $\mathbf{s}$ obeying $\mathbf{c}^T\mathbf{s} = \gamma$ where $\mathbf{c}$ is a vector and γ is a scalar.[3]
7. The interior of a circular disk in 2-space; *i.e.*, points (x, y) obeying

$$(x-a)^2 + (y-b)^2 < c^2$$

for real a, b and c.

[2] $[a, b]$ means the set of numbers x such that $a \leq x \leq b$.
[3] T used in a superscript means to take the transpose of a vector or a matrix.

We note that $\mathbf{R}_+^n$ (example 3) defines the set of n-dimensional block slowness models such that the slowness of each cell is a positive number. $C_+(\mathbf{R}^3)$ (example 4) is the space of positive, continuous 3-D slowness distributions.

PROPOSITION 2.2.1 *If $\mathcal{A}_1$ and $\mathcal{A}_2$ are convex sets, then $\mathcal{A}_1 \cap \mathcal{A}_2$ is a convex set.*[4]

Proof: If $\mathcal{A}_1 \cap \mathcal{A}_2$ is empty, it is convex by default (one cannot find $\mathbf{s}_1$, $\mathbf{s}_2$ and λ which disobey the definition).

Assume the intersection is not empty and let $\mathbf{s} = \lambda\mathbf{s}_1 + (1-\lambda)\mathbf{s}_2$ for some $0 \le \lambda \le 1$ and $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{A}_1 \cap \mathcal{A}_2$. Since $\mathcal{A}_1$ and $\mathcal{A}_2$ are each convex, we must have $\mathbf{s} \in \mathcal{A}_1$ and $\mathbf{s} \in \mathcal{A}_2$. Consequently, $\mathbf{s} \in \mathcal{A}_1 \cap \mathcal{A}_2$. ▮

DEFINITION 2.2.2 (CONE) *A set $\mathcal{A} \subseteq \mathcal{S}$ is a cone if, for every $\mathbf{s} \in \mathcal{A}$ and every number $\gamma > 0$, we have $\gamma\mathbf{s} \in \mathcal{A}$.*

Examples 1–4 of convex sets given above are also examples of cones. We infer that the set of positive slowness models (block or continuous) is convex and conical (a *convex cone*).

DEFINITION 2.2.3 (LINEAR FUNCTIONAL) *The functional $f\colon \mathcal{S} \to \mathbf{R}$ is linear if, for all $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$ and real numbers λ_1, λ_2, we have*[5]

$$f(\lambda_1\mathbf{s}_1 + \lambda_2\mathbf{s}_2) = \lambda_1 f(\mathbf{s}_1) + \lambda_2 f(\mathbf{s}_2). \tag{2.3}$$

Considering $\lambda_1 = \lambda_2 = 0$, note that a linear functional necessarily vanishes at the origin. We will also need to consider the broader class of functionals that are linear except for a shift at the origin.

DEFINITION 2.2.4 (SHIFTED LINEAR FUNCTIONAL) *The functional $f\colon \mathcal{S} \to \mathbf{R}$ is shifted linear if the functional*

$$g(\mathbf{s}) \equiv f(\mathbf{s}) - f(0) \tag{2.4}$$

is linear.

DEFINITION 2.2.5 (CONVEX FUNCTIONAL) *Let $\mathcal{A}$ be a convex set in $\mathcal{S}$. A functional $f\colon \mathcal{A} \to \mathbf{R}$ is convex if, for every $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{A}$ and number $\lambda \in [0,1]$, we have*

$$f(\lambda\mathbf{s}_1 + (1-\lambda)\mathbf{s}_2) \le \lambda f(\mathbf{s}_1) + (1-\lambda) f(\mathbf{s}_2). \tag{2.5}$$

DEFINITION 2.2.6 (CONCAVE FUNCTIONAL) *A functional f is concave if $(-f)$ is convex.*

DEFINITION 2.2.7 (HOMOGENEOUS FUNCTIONAL) *Let $\mathcal{A}$ be a cone in $\mathcal{S}$. A functional $f\colon \mathcal{A} \to \mathbf{R}$ is homogeneous if, for every $\mathbf{s} \in \mathcal{A}$ and $\gamma > 0$, we have*

$$f(\gamma\mathbf{s}) = \gamma f(\mathbf{s}). \tag{2.6}$$

[4] $\mathcal{A}_1 \cap \mathcal{A}_2$ denotes the *intersection* of sets $\mathcal{A}_1$ and $\mathcal{A}_2$ (i.e., the set of elements common to both sets).

[5] $f\colon \mathcal{S} \to \mathbf{R}$ means: the function f which maps each element of the set $\mathcal{S}$ to an element of the set $\mathbf{R}$.

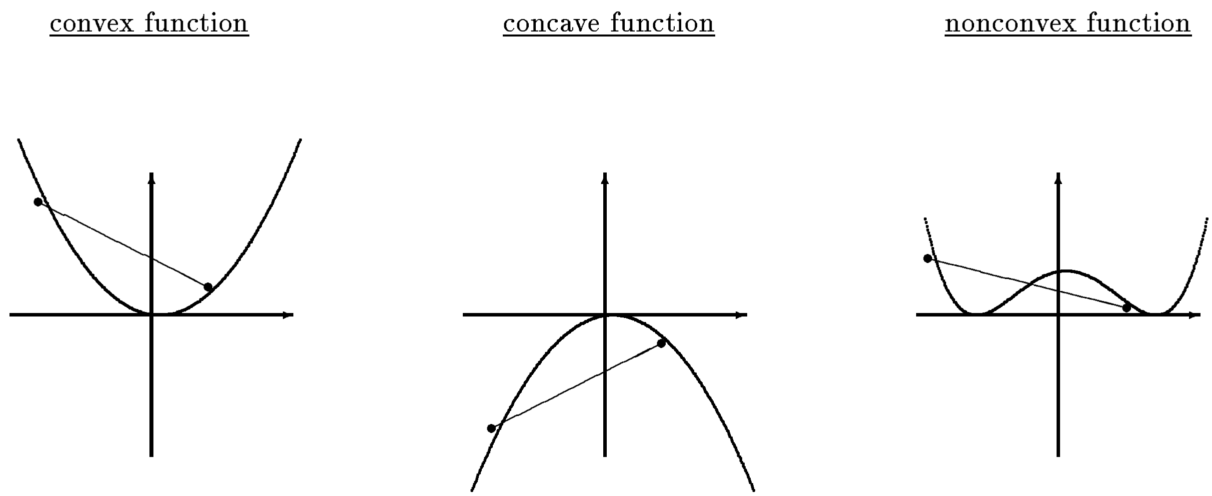

Figure 2.3: Examples of convex, concave, and nonconvex functionals.

It should be clear that every linear functional is also convex, concave, and homogeneous.

Problems

PROBLEM 2.2.1 *Is the union of two convex sets convex? Give an example.*

PROBLEM 2.2.2 *Decide whether the following sets are convex:*

1. *the interior of a cube;*
2. *the interior of a tetrahedron;*
3. *the interior of a rectangular prism;*
4. *any compact region in n-dimensional vector space, all of whose boundaries are hyperplanes;*
5. *the interior of an ellipsoid;*
6. *the interior of an n-dimensional sphere;*
7. *the interior of two partially overlapping spheres;*
8. *the interior of a boomerang;*
9. *the cheesy part of a swiss cheese;*
10. *the interior of any object having a rough surface.*

PROBLEM 2.2.3 *If* $\mathbf{c}$ *is an arbitrary* n*-vector, which of the following functionals is linear in* $\mathbf{s}$*?*

1. $\mathbf{c}^T\mathbf{s}$;
2. $\mathbf{s}\mathbf{s}^T\mathbf{c}$;
3. $(\mathbf{s}-\mathbf{c})^T(\mathbf{s}-\mathbf{c})$.

PROBLEM 2.2.4 *Show that a linear functional is convex, concave, and homogeneous.*

PROBLEM 2.2.5 *Is a shifted linear functional convex, concave, and/or homogeneous?*

PROBLEM 2.2.6 *Are all cones convex? If not, give an example of a nonconvex cone.*

2.3 Properties of Traveltime Functionals

PROPOSITION 2.3.1 τ^P *is a linear functional.*

The proof of this stems from the fact that integration is a linear functional of the integrand. Since it is linear, it follows that τ^P is also convex, concave, and homogeneous.

PROPOSITION 2.3.2 τ^* *is a homogeneous functional.*

Proof: Given $\gamma > 0$ we have

$$\tau^*(\gamma\mathbf{s}) = \min_P \tau^P(\gamma\mathbf{s}). \tag{2.7}$$

Using the linearity of τ^P,

$$\tau^*(\gamma\mathbf{s}) = \min_P \gamma\tau^P(\mathbf{s}) = \gamma \min_P \tau^P(\mathbf{s}) = \gamma\tau^*(\mathbf{s}). \quad \blacksquare \tag{2.8}$$

PROPOSITION 2.3.3 τ^* *is a concave functional.*

Proof: Given slowness models $\mathbf{s}_1$ and $\mathbf{s}_2$ and $\lambda \in [0,1]$, let $\mathbf{s} = \lambda\mathbf{s}_1 + (1-\lambda)\mathbf{s}_2$. Letting $P^*(\mathbf{s})$ be the Fermat ray path for $\mathbf{s}$, we have

$$\tau^*(\mathbf{s}) = \tau^{P^*(\mathbf{s})}(\mathbf{s}). \tag{2.9}$$

The linearity of τ^P then implies

$$\tau^*(\mathbf{s}) = \lambda\tau^{P^*(\mathbf{s})}(\mathbf{s}_1) + (1-\lambda)\tau^{P^*(\mathbf{s})}(\mathbf{s}_2). \tag{2.10}$$

Since τ^* minimizes τ^P for any fixed model, it must be the case that $\tau^{P^*(\mathbf{s})}(\mathbf{s}_1) \geq \tau^*(\mathbf{s}_1)$ and similarly for $\mathbf{s}_2$. Further, λ and $(1-\lambda)$ are non-negative. Therefore, (2.10) implies

$$\tau^*(\mathbf{s}) \geq \lambda\tau^*(\mathbf{s}_1) + (1-\lambda)\tau^*(\mathbf{s}_2). \quad \blacksquare \tag{2.11}$$

2.4 Feasibility Sets

Given the set of observed traveltimes, t_i for $i = 1, \ldots, m$, we define two sets of models:

DEFINITION 2.4.1 (LOCAL FEASIBILITY SET) *The local feasibility set with respect to a set of trial ray paths $\mathcal{P} = \{P_1, \ldots, P_m\}$ and observed traveltimes $t_1, \ldots, t_m$ is*

$$\mathcal{F}^{\mathcal{P}} = \{\mathbf{s} \mid \tau_i^P(\mathbf{s}) \geq t_i, \quad \textit{for all} \quad i = 1, \ldots, m\}. \tag{2.12}$$

DEFINITION 2.4.2 (GLOBAL FEASIBILITY SET) *The global feasibility set with respect to the observed traveltimes $t_1, \ldots, t_m$ is*

$$\mathcal{F}^* = \{\mathbf{s} \mid \tau_i^*(\mathbf{s}) \geq t_i, \quad \textit{for all} \quad i = 1, \ldots, m\}. \tag{2.13}$$

Now we show that the concavity of τ_i^P and τ_i^* implies the convexity of $\mathcal{F}^{\mathcal{P}}$ and $\mathcal{F}^*$.

THEOREM 2.4.1 *$\mathcal{F}^{\mathcal{P}}$ is a convex set.*

Proof: Suppose $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{F}^{\mathcal{P}}$ and let $\mathbf{s}_\lambda = \lambda \mathbf{s}_1 + (1 - \lambda)\mathbf{s}_2$ where $0 \leq \lambda \leq 1$. Since, for each i, τ_i^P is a concave (actually linear) functional, we have

$$\tau_i^P(\mathbf{s}_\lambda) \geq \lambda \tau_i^P(\mathbf{s}_1) + (1 - \lambda)\tau_i^P(\mathbf{s}_2). \tag{2.14}$$

(Although equality applies in the present case, the "greater than or equal to" is important in the next proof.) But $\tau_i^P(\mathbf{s}_1), \tau_i^P(\mathbf{s}_2) \geq t_i$ and λ and $(1 - \lambda)$ are non-negative. Therefore,

$$\tau_i^P(\mathbf{s}_\lambda) \geq \lambda t_i + (1 - \lambda)t_i = t_i. \tag{2.15}$$

Thus, $\mathbf{s}_\lambda \in \mathcal{F}^{\mathcal{P}}$. ∎

THEOREM 2.4.2 *$\mathcal{F}^*$ is a convex set.*

The proof proceeds in analogy with the previous proof, with τ_i^* replacing τ_i^P, but the inequalities come into play this time.

The next theorem follows easily from an analysis of Figure 2.1.

THEOREM 2.4.3 *Given any model $\mathbf{s}$, there exists a finite scalalr $\gamma^* > 0$ such that $\gamma \mathbf{s} \in \mathcal{F}^*$ for all $\gamma \geq \gamma^*$.*

Proof: Let

$$\gamma^* = \max_{k \in \{1, \ldots, m\}} \frac{t_k}{\tau_k^*(\mathbf{s})}. \tag{2.16}$$

For any i, τ_i^* is homogeneous, implying

$$\tau_i^*(\gamma^* \mathbf{s}) = \gamma^* \tau_i^*(\mathbf{s}) = \tau_i^*(\mathbf{s}) \max_k \frac{t_k}{\tau_k^*(\mathbf{s})} \geq \tau_i^*(\mathbf{s}) \frac{t_i}{\tau_i^*(\mathbf{s})} = t_i. \tag{2.17}$$

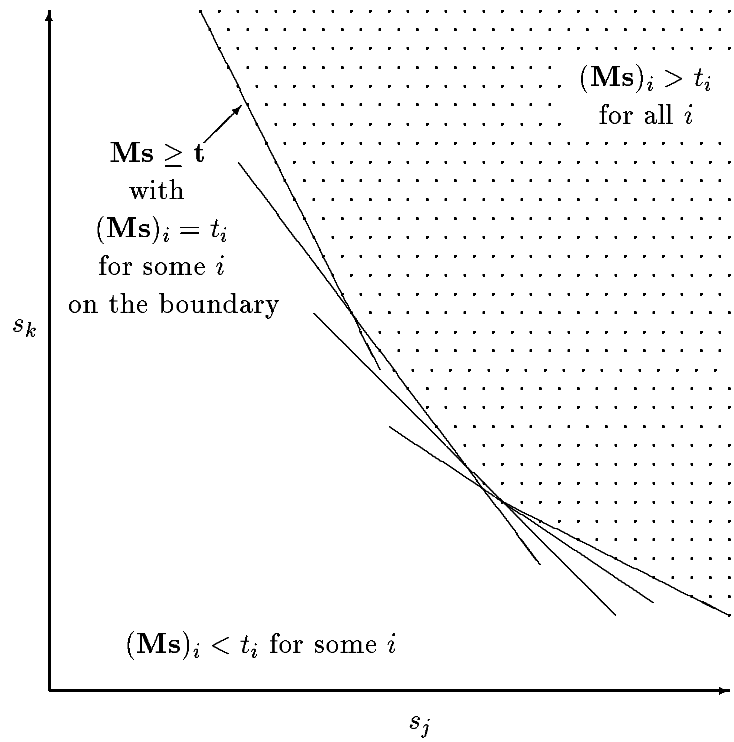

Figure 2.4: The defining conditions for the feasible and infeasible parts of the model space and the boundary separating them.

We see that $\gamma^*\mathbf{s}$ satisfies all the feasibility constraints, so it is in $\mathcal{F}^*$, and so is $\gamma\mathbf{s}$ for any $\gamma > \gamma^*$. ▮

We can decompose $\mathcal{F}^*$ into two parts: its *boundary* and its *interior*. The boundary of $\mathcal{F}^*$, denoted $\partial\mathcal{F}^*$, comprises feasible models $\mathbf{s}$ which satisfy some feasibility constraint with equality, *i.e.*,

$$\partial\mathcal{F}^* = \{\mathbf{s} \in \mathcal{F}^* \mid \tau_i^*(\mathbf{s}) = t_i, \text{ for some } i\}. \tag{2.18}$$

Models in the interior of $\mathcal{F}^*$, denoted $\operatorname{Int}\mathcal{F}^* = \mathcal{F}^* - \partial\mathcal{F}^*$, satisfy all constraints with inequality:

$$\operatorname{Int}\mathcal{F}^* = \{\mathbf{s} \in \mathcal{F}^* \mid \tau_i^*(\mathbf{s}) > t_i, \text{ for all } i\}. \tag{2.19}$$

These characteristics of the feasible set are illustrated in Figure 2.4.

Problems

PROBLEM 2.4.1 *Prove that $\mathcal{F}^*$ is convex.*

PROBLEM 2.4.2 *Define another type of feasibility set $\mathcal{R}$ by*

$$\mathcal{R} = \{\mathbf{s} \mid \sum_{i=1}^{m} \tau_i^*(\mathbf{s}) \geq \sum_{i=1}^{m} t_i\}.$$

Is $\mathcal{R}$ convex? What is the relationship between this set and the other ones defined in this section? (Larger or smaller set?) [relaxed constraints]

PROBLEM 2.4.3 *Consider the velocity vector space related to the slowness vector space through the nonlinear transform $\mathbf{v} = (1/s_1, 1/s_2, \ldots, 1/s_n)$. Suppose there is some reason to expect that all models solving our inversion problem should lie in a part of the velocity vector space satisfying the hyperplane constraint given by*

$$\mathcal{V} = \{\mathbf{v} \mid \mathbf{d}^T\mathbf{v} \geq \beta\},$$

where β is some positive constant and $\mathbf{d}$ is a nonnegative vector. Now, let $\mathcal{V}'$ be the set of slownesses corresponding to

$$\mathcal{V}' = \{\mathbf{s} \mid \mathbf{v} \in \mathcal{V}\}$$

and define the slowness overlap set

$$\mathcal{D} = \mathcal{V}' \cap \mathcal{F}^*.$$

Assuming that the set $\mathcal{D}$ is not empty, is it convex?

2.5 Convex Programming for Inversion

We first define convex programming for first-arrival traveltime inversion. Then we present some basic theorems about convex programming in this context.

DEFINITION 2.5.1 *Let $\Phi(\mathbf{s})$ be any convex functional of $\mathbf{s}$. Then the convex nonlinear programming problem associated with Φ is to minimize $\Phi(\mathbf{s})$ subject to the global feasibility constraints $\tau_i^*(\mathbf{s}) \geq t_i$, for $i = 1, \ldots, m$.*

DEFINITION 2.5.2 *Let*

$$\Psi^{\mathcal{P}}(\mathbf{s}) = \sum_{i=1}^{m} w_i[\tau_i^{P_i}(\mathbf{s}) - t_i]^2 \tag{2.20}$$

for some positive weights $\{w_i\}$ and some set of ray paths $\mathcal{P} = \{P_1, \ldots, P_m\}$. Then, the convex linear programming problem associated with $\Psi^{\mathcal{P}}$ is to minimize $\Psi^{\mathcal{P}}(\mathbf{s})$ subject to the local feasibility constraints $\tau_i^{P_i}(\mathbf{s}) \geq t_i$, for $i = 1, \ldots, m$.

THEOREM 2.5.1 *Every local minimum $\mathbf{s}^*$ of the convex nonlinear programming problem associated with $\Phi(\mathbf{s})$ is a global minimum.*

THEOREM 2.5.2 *Every local minimum* $\mathbf{s}^*$ *of the convex linear programming problem associated with* $\Psi^{\mathcal{P}}(\mathbf{s})$ *is a global minimum.*

Proof: This proof follows one given by Fiacco and McCormick [1990]. Let $\mathbf{s}^*$ be a local minimum. Then, by definition, there is a compact set $\mathcal{C}$ such that $\mathbf{s}^*$ is in the interior of $\mathcal{C} \cap \mathcal{F}^*$ and

$$\Phi(\mathbf{s}^*) = \min_{\mathcal{C}\cap\mathcal{F}^*} \Phi(\mathbf{s}). \tag{2.21}$$

If $\mathbf{s}$ is any point in the feasible set $\mathcal{F}^*$ and $0 \leq \lambda \leq 1$ such that $\mathbf{s}^*_\lambda \equiv \lambda\mathbf{s}^* + (1-\lambda)\mathbf{s}$ is in $\mathcal{C} \cap \mathcal{F}^*$, then

$$\Phi(\mathbf{s}) \geq \frac{\Phi(\mathbf{s}^*_\lambda) - \lambda\Phi(\mathbf{s}^*)}{1-\lambda} \geq \frac{\Phi(\mathbf{s}^*) - \lambda\Phi(\mathbf{s}^*)}{1-\lambda} = \Phi(\mathbf{s}^*). \tag{2.22}$$

The first step of (2.22) follows from the convexity of Φ and the second from the fact that $\mathbf{s}^*$ is a minimum in $\mathcal{C} \cap \mathcal{F}^*$. Convexity of $\mathcal{F}^*$ guarantees that the convex combination $\mathbf{s}^*_\lambda$ lies in the feasible set. This completes the proof of the first theorem.

The proof of the second theorem follows that of the first once we have shown that the function $\Psi^{\mathcal{P}}$ is convex. Consider a term of $\Psi^{\mathcal{P}}$

$$\begin{aligned}
[\tau_i^{P_i}(\lambda\mathbf{s}_1 + (1-\lambda)\mathbf{s}_2) - t_i]^2 &= [\lambda\tau_i^{P_i}(\mathbf{s}_1) + (1-\lambda)\tau_i^{P_i}(\mathbf{s}_2) - t_i]^2 \\
&= \lambda[\tau_i^{P_i}(\mathbf{s}_1) - t_i]^2 + (1-\lambda)[\tau_i^{P_i}(\mathbf{s}_2) - t_i]^2 \\
&\quad -\lambda(1-\lambda)[\tau_i^{P_i}(\mathbf{s}_1) - \tau_i^{P_i}(\mathbf{s}_2)]^2 \\
&\leq \lambda[\tau_i^{P_i}(\mathbf{s}_1) - t_i]^2 + (1-\lambda)[\tau_i^{P_i}(\mathbf{s}_2) - t_i]^2.
\end{aligned}$$

Then, if $\mathbf{s}_\lambda = \lambda\mathbf{s}_1 + (1-\lambda)\mathbf{s}_2$,

$$\Psi^{\mathcal{P}}(\mathbf{s}_\lambda) \leq \lambda\Psi^{\mathcal{P}}(\mathbf{s}_1) + (1-\lambda)\Psi^{\mathcal{P}}(\mathbf{s}_2), \tag{2.23}$$

so $\Psi^{\mathcal{P}}$ is a convex function. ▮

Thus, linear inversion is a convex programming problem. These results show further that, if we could find a convex functional of slowness $\mathbf{s}$ pertinent to the nonlinear inversion problem, then the nonlinear programming problem would be easy (*i.e.*, proofs of convergence become trivial), because there would be no local minima. However, this analysis does not guarantee the existence of such a functional, nor do we know how to construct such a functional even if we suppose one exists. It remains an open question whether an appropriate convex functional for nonlinear seismic inversion can be found.

I expect this question to remain open for a long time, but nevertheless challenge the reader to prove me wrong in this prediction.