Universität Stuttgart
Mathematisches Institut A
1. Lehrstuhl

Angewandte Analysis
Pfaffenwaldring 57
D–70569 Stuttgart

http://www.mathematik.uni-stuttgart.de/mathA/lst1/

Finite elastoplasticity, Lie groups and geodesics on SL(d)

Alexander Mielke

Mathematisches Institut A, Universität Stuttgart, D-70569 Stuttgart

mielke@mathematik.uni-stuttgart.de

21 September 2000

Abstract

The notions of nonlinear plasticity with finite deformations is intepreted in the sense of Lie groups. In particular, the plastic tensor $\boldsymbol{P} = \boldsymbol{F}_{\mathrm{p}}^{-1}$ is considered as element of the Lie group SL(d). Moreover, the plastic dissipation defines a left–invariant Finsler metric on the tangent bundle of this Lie group. In the case of single crystal plasticity this metric is given in terms of the different slip systems and is piecewise affine on each tangent space. For von Mises plasticity the metric is a left–invariant Riemannian metric. A main goal is to study the associated distance metric and the geodesic curves.

Contents

1 Introduction

We give a formulation of perfect plasticity in the context of finite strain. This formulation is based on two functionals, namely the stored–energy density ψ (also called the Helmholtz free engergy) and the dissipation functional $\widehat{\Delta}$. The deformation of a body $\Omega \subset \mathbb{R}^d$ is a mapping $\boldsymbol{\varphi} : \Omega \to \mathbb{R}^d$; and the deformation gradient $\mathrm{D}\boldsymbol{\varphi} \in \mathbb{R}^{d\times d}$ is denoted by $\boldsymbol{F}$. In perfect finite elastoplasticity the function ψ depends on the deformation gradient $\boldsymbol{F}$ and the plastic variable $\boldsymbol{P} = \boldsymbol{F}_{\mathrm{p}}^{-1}$ solely via the elastic part $\boldsymbol{F}_{\mathrm{e}}$ of the deformation gradient, i.e. $\boldsymbol{F}_{\mathrm{e}} = \boldsymbol{F}\boldsymbol{P} = \boldsymbol{F}\boldsymbol{F}_{\mathrm{p}}^{-1}$. We write $\psi = \widehat{\psi}(\boldsymbol{F},\boldsymbol{P}) = \Psi(\boldsymbol{F}\boldsymbol{P})$. This multiplicative split was introduced in [Lee69] and its use is nowadays well–established, see e.g. [SiO85, Sim88, MiS92, Hac97, Gur00].

The main point in this paper is to work out the Lie group structure which is central to finite plasticity. For the usage of Lie group theory in mechanics we refer to [AbM78, MaH83]. In this work all appearing Lie groups are matrix groups, so the paper can be read without previous knowledge in Lie group theory. The deformation gradient $\boldsymbol{F}$ is considered as an element of the general linear group

$$\mathrm{GL}(d) = \mathrm{GL}_+(d,\mathbb{R}) = \{\boldsymbol{F} \in \mathbb{R}^{d\times d} : \det \boldsymbol{F} > 0\},$$

and the plastic variable $\boldsymbol{P}$ is considered as an element of a Lie subgroup $G \subset \mathrm{GL}(d)$; in many cases G is the special linear group

$$\mathrm{SL}(d) = \{\boldsymbol{P} \in \mathrm{GL}(d) : \det \boldsymbol{P} = 1\}.$$

The dissipation functional $\widehat{\Delta}$ is a mapping from $\mathrm{T}G$ into $[0,\infty)$ such that

$$\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) = \Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}) \quad \text{with} \quad \Delta(\alpha\boldsymbol{\xi}) = \alpha\Delta(\boldsymbol{\xi})$$

for all $(\boldsymbol{P},\dot{\boldsymbol{P}}) \in \mathrm{T}G$ and $\alpha \geq 0$. The first condition expresses the fact that previous plastic deformations do not influence the dissipation (i.e. for each $\boldsymbol{G} \in G$ the dissipation of the process $t \mapsto \boldsymbol{G}\boldsymbol{P}(t)$ is the same). The second condition gives rate independency of the model. The function Δ is defined on the Lie algebra $\boldsymbol{g} = \mathrm{T}_{\boldsymbol{I}}G$ and is assumed to be convex and coercive, i.e., $r_1\|\boldsymbol{\xi}\| \leq \Delta(\boldsymbol{\xi}) \leq r_2\|\boldsymbol{\xi}\|$.

The two functionals $\widehat{\psi}$ and $\widehat{\Delta}$ (or equivalently Ψ and Δ) together with the principle of maximal dissipation (cf. [SiO85, ZiW87, Sim88]) define the full equations of elastoplasticity. We give here an alternate form using the subgradient $\partial\Delta(\boldsymbol{\xi})$:

$$\begin{aligned}
&\boldsymbol{\Sigma} = \tfrac{\partial}{\partial \boldsymbol{F}}\widehat{\psi}(\boldsymbol{F},\boldsymbol{P}) = \mathrm{D}\Psi(\boldsymbol{F}\boldsymbol{P})\boldsymbol{P}^\top,\\
&\boldsymbol{0} \in \partial\Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}) + \boldsymbol{P}^\top \tfrac{\partial}{\partial \boldsymbol{P}}\widehat{\psi}(\boldsymbol{F},\boldsymbol{P}) = \partial\Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}) + (\boldsymbol{F}\boldsymbol{P})^\top \mathrm{D}\Psi(\boldsymbol{F}\boldsymbol{P}).
\end{aligned} \tag{1.1}$$

The second equation is posed on $\boldsymbol{g}$. Together with the relation $\boldsymbol{F} = \mathrm{D}\boldsymbol{\varphi}$ and the elastic equilibrium condition $-\mathrm{div}\boldsymbol{\Sigma} = \boldsymbol{f}_{\mathrm{ext}}$ and suitable boundary conditions this defines the classical plastic flow rule. The equivalenve between these equations an the classical mechanical flow rules involving the yield surface is given in Section 2.

As in [CHM00] we propose an integrated version of (1.1) which implies (1.1) but not vice versa. To this end we need to define a non–symmetric distance metric on G as follows:

$$\begin{aligned}
&\widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_1) = \inf\{I_\Delta(\boldsymbol{P}(\cdot)) : \boldsymbol{P} \in \mathrm{C}^1([0,1],G), \boldsymbol{P}(0) = \boldsymbol{P}_0, \boldsymbol{P}(1) = \boldsymbol{P}_1\}\\
&\text{with } I_\Delta(\boldsymbol{P}(\cdot)) = \textstyle\int_0^1 \Delta(\boldsymbol{P}(t)^{-1}\dot{\boldsymbol{P}}(t))\,\mathrm{d}t.
\end{aligned} \tag{1.2}$$

The distance metric $\widehat{D}$ is a so–called Finsler–Minkowsky metric generated by the infinitesimal metric $\widehat{\Delta} : \mathrm{T}G \to [0,\infty); (\boldsymbol{G},\dot{\boldsymbol{G}}) \mapsto \Delta(\boldsymbol{G}^{-1}\dot{\boldsymbol{G}})$. The metric is invariant with respect to the left translation. In the special situation where $\Delta(\boldsymbol{\xi}) = \langle \mathbb{B}\boldsymbol{\xi},\boldsymbol{\xi}\rangle^{1/2}$ with $\mathbb{B} = \mathbb{B}^* > 0$ we obtain a Riemannian metric.

By integrating $\widehat{\psi}$ and $\widehat{\Delta}$ over the body $\Omega \subset \mathbb{R}^d$ we define the nonlinear functionals

$$\begin{aligned}\mathcal{E}(t,\boldsymbol{\varphi},\boldsymbol{P}) &= \textstyle\int_\Omega \{\Psi([\mathrm{D}\boldsymbol{\varphi}]\boldsymbol{P}) - \boldsymbol{f}_{\mathrm{ext}}(t,\cdot)\cdot\boldsymbol{\varphi}\}\,\mathrm{d}x - \int_{\Gamma_{\mathrm{tract}}} \boldsymbol{t}_{\mathrm{ext}}(t,\cdot)\cdot\boldsymbol{\varphi}\,\mathrm{d}a,\\ \widehat{D}_\Omega(\boldsymbol{P}_0,\boldsymbol{P}_1) &= \textstyle\int_\Omega \widehat{D}(\boldsymbol{P}_0(x),\boldsymbol{P}_1(x))\,\mathrm{d}x.\end{aligned}$$

Here, $\mathcal{E}$ is the elastic and potential energy stored in the body under given $\boldsymbol{\varphi} : \Omega \to \mathbb{R}^d$ and $\boldsymbol{P} : \Omega \to G$. The functional $\widehat{D}_\Omega$ gives the minimal dissipation arising from a change of the plastic variable from $\boldsymbol{P}_0 : \Omega \to G$ to the new state $\boldsymbol{P}_1 : \Omega \to G$.

The global nondifferentiated form of the plastic flow problem now takes the form

$$\begin{array}{ll}\textbf{global stability:} & \mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) \leq \mathcal{E}(t,\boldsymbol{\varphi}^*,\boldsymbol{P}^*) + \widehat{D}_\Omega(\boldsymbol{P}(t),\boldsymbol{P}^*)\\ & \text{for all } (\boldsymbol{\varphi}^*,\boldsymbol{P}^*) \text{ with } \boldsymbol{\varphi}^*|_{\Gamma_{\mathrm{displ}}} = \boldsymbol{\varphi}_0(t,\cdot);\\ \textbf{energy balance:} & \mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) + \int_s^t\int_\Omega \Delta(\boldsymbol{P}(\tau,x)^{-1}\dot{\boldsymbol{P}}(\tau,x))\,\mathrm{d}x\,\mathrm{d}\tau\\ & = \mathcal{E}(s,\boldsymbol{\varphi}(s),\boldsymbol{P}(s)) + \int_s^t \frac{\partial}{\partial t}\mathcal{E}(\tau,\boldsymbol{\varphi}(\tau),\boldsymbol{P}(\tau))\,\mathrm{d}\tau.\end{array} \tag{1.3}$$

A similar energetic formulation for phase transformation problems was given in [MiT99, MTL00, GoM00].

Such a formulation has the major advantage, that it immediately suggests an incremental formulation which has a variational form. If $0 = t_0 < t_1 < \ldots t_N = T$ is a time discretization and $\boldsymbol{P}_0 : \Omega \to G$ the initial value, then the time stepping algorithm reads

> **Incremental Problem**: Given $(\boldsymbol{\varphi}_k,\boldsymbol{P}_k)$ at time t_k,
> find $(\boldsymbol{\varphi}_{k+1},\boldsymbol{P}_{k+1})$ such that it minimizes the
> function $\mathrm{T}_{k+1} : (\boldsymbol{\varphi},\boldsymbol{P}) \mapsto \mathcal{E}(t_{k+1},\boldsymbol{\varphi},\boldsymbol{P}) + \widehat{D}_\Omega(\boldsymbol{P}_k,\boldsymbol{P})$.

In Lemma 2.1 we show that all solutions of this incremental problem satisfy global stability as well as a discretized version of the energy balance in (1.3). Incremental formulations for rate–independent processes which solely depend on minimization are well–known in abstract sweeping processes (cf. [Mon93]) and in linearized elastoplasticity, cf. e.g. [HaR95]. The usage of such formulations in finite elastoplasticity is only recent [OrR99, CHM00].

The above theory generates a need of a characterization of the Finsler–Minkowsky distance metric $\widehat{D} : G\times G \to [0,\infty)$ as defined in (1.2). By left invariance (i.e. $\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) = \Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}})$) we have

$$\widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_1) = D(\boldsymbol{P}_0^{-1}\boldsymbol{P}_1) \quad \text{with } D(\boldsymbol{P}) := \widehat{D}(\boldsymbol{I},\boldsymbol{P}).$$

Recalling $\Delta(\boldsymbol{\xi}) \leq r_2\|\boldsymbol{\xi}\|$ we immediately obtain an upper bound $D(\boldsymbol{P}) \leq r_2\|\log\boldsymbol{P}\|$ whenever a logarithm $\boldsymbol{\xi} \in \boldsymbol{g}$ with $\boldsymbol{P} = \mathrm{e}^{\boldsymbol{\xi}}$ exists. This already indicates that the Lie algebraic exponential map does not provide the geodesic curves, i.e., the curves $t \mapsto \boldsymbol{P}(t) = \mathrm{e}^{t\boldsymbol{\xi}}$ are not minimizers in (1.2) in general. In Section 6 we provide a geodesic exponential map which is surjective, see also M in (1.5) below.

Sections 4 to 6 are concerned only with the study of the metric $\widehat{D}$ and its shortest paths. In Theorem 5.3 we show that the infimum in (1.2) is always attained, i.e., there is always a curve $\boldsymbol{P} \in \mathrm{C}^{\mathrm{Lip}}([0,1], G)$ with $\boldsymbol{P}(0) = \boldsymbol{P}_0$ and $\boldsymbol{P}(1) = \boldsymbol{P}_1$ such that $\widehat{D}(\boldsymbol{P}_0, \boldsymbol{P}_1) = I_\Delta(\boldsymbol{P}(\cdot))$. A geodesic curve $\boldsymbol{P} : (t_0, t_1) \to G$ is a curve which is locally a shortest path.

The difficulty for Finsler metrics arises from two facts: first, $\Delta : \boldsymbol{g} \setminus \{\boldsymbol{0}\} \to [0, \infty)$ may be non differentiable and, second, the subgradient $\mathbb{Q} = \partial\Delta(\boldsymbol{0}) \subset \boldsymbol{g}^*$ at 0 may be not strictly convex, where

$$\mathbb{Q} = \partial\Delta(\boldsymbol{0}) = \{\boldsymbol{\eta} \in \boldsymbol{g}^* : \langle \boldsymbol{\eta}, \boldsymbol{\xi} \rangle \leq \Delta(\boldsymbol{\xi}) \text{ for all } \boldsymbol{\xi}\}.$$

In this general situation we do not have a full characterization of shortest paths or geodesic curves. However, we derive a nonsmooth counterpart to Noether's theorem. For instance, from the left invariance of $\widehat{\Delta}$ it follows that for each shortest path $\boldsymbol{P}$ there exists an $\boldsymbol{\eta} \in \boldsymbol{g}^*$ such that

$$\boldsymbol{P}^\top \boldsymbol{\eta} \boldsymbol{P}^{-\top} \in \partial\Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}) \tag{1.4}$$

holds almost everywhere along the curve. We continue to use the notation $(\,\cdot\,)^\top$ for the transposed matrix where throughout the Euclidean scalar product in $\mathbb{R}^d$ is used. See Theorem 5.4 for the exact statement in Lie group notation. Examples of nonsmooth shortest paths and geodesics are given there as well.

In Section 6 we study von Mises plasticity where Δ generates a left invariant Riemann metric on $G = \mathrm{SL}(d)$ which is additionally right invariant with respect to $\mathrm{SO}(d)$. The latter invariance corresponds to isotropy of the elastoplastic model. Starting from

$$\Delta(\boldsymbol{\xi}) = (\gamma_2 \|\boldsymbol{\xi}_{\mathrm{sym}}\|^2 + \gamma_3 \|\boldsymbol{\xi}_{\mathrm{anti}}\|^2)^{1/2} \quad \text{where } \boldsymbol{\xi}_{\mathrm{sym}} = \tfrac{1}{2}(\boldsymbol{\xi} + \boldsymbol{\xi}^\top),\ \boldsymbol{\xi}_{\mathrm{anti}} = \tfrac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\xi}^\top)$$

with $\gamma_2, \gamma_3 > 0$ we find that all geodesics are smooth and have the form

$$\boldsymbol{P}(t) = M(t(\boldsymbol{\alpha} + \boldsymbol{\omega})) := \boldsymbol{P}(0) \mathrm{e}^{t\boldsymbol{\alpha}} \mathrm{e}^{t\boldsymbol{\omega}} \text{ with } \boldsymbol{\alpha} = \tfrac{1}{\gamma_2} \boldsymbol{\eta}^\top \text{ and } \boldsymbol{\omega} = (\tfrac{1}{\gamma_2} + \tfrac{1}{\gamma_3}) \boldsymbol{\eta}_{\mathrm{anti}}. \tag{1.5}$$

Here $\boldsymbol{\eta}$ is defined such that (1.4) holds. This leads to the formula

$$D(\boldsymbol{P}) = \widehat{D}(\boldsymbol{I}, \boldsymbol{P}) = \min\{(\gamma_2 \|\boldsymbol{\xi}_{\mathrm{sym}}\|^2 + \gamma_3 \|\boldsymbol{\xi}_{\mathrm{anti}}\|^2)^{1/2} : \boldsymbol{P} = \mathrm{e}^{\boldsymbol{\xi}_{\mathrm{sym}} - \delta \boldsymbol{\xi}_{\mathrm{anti}}} \mathrm{e}^{(1+\delta)\boldsymbol{\xi}_{\mathrm{anti}}}\}$$

where $\delta = \gamma_3/\gamma_2$.

Acknowledgments. The author is grateful to Dirk Mittenhuber for fruitful e–mail exchanges. The research was partially supported by DFG within the SFB 404.

2 Formulations of elastoplasticity

The aim of this section is to motivate the model for elastoplasticity at finite deformations from the mechanical point of view, for more background we refer to [Sim88, Hac97, Gur00]. Then we reformulate the mechanical notion by using methods from convex analysis based on subgradients. This leads to a formulation involving energy balance and local (infinitesimal) stability. We will then argue that local stability can be replaced by global stability which then justifies the incremental problem in variational form.

We consider perfect finite plasticity, i.e., we do not take into account plastic hardening effects. The body is given by a domain $\Omega \subset \mathbb{R}^d$, where the dimension d is usually equal to 3 but $d \in \{1, 2\}$ is also relevant in certain special situations. The deformation is $\boldsymbol{\varphi} : \Omega \to \mathbb{R}^d$ and its gradient $\boldsymbol{F} = \mathrm{D}\boldsymbol{\varphi}$ is decomposed multiplicatively into the elastic part $\boldsymbol{F}_{\mathrm{e}}$ and the plastic part $\boldsymbol{F}_{\mathrm{p}}$

$$\mathrm{D}\boldsymbol{\varphi} = \boldsymbol{F} = \boldsymbol{F}_{\mathrm{e}}\boldsymbol{F}_{\mathrm{p}}. \tag{2.1}$$

The plastic part $\boldsymbol{F}_{\mathrm{p}}$ is not a gradient. It corresponds to the atomic disarrangement in single crystals arising through slips. In polycrystals $\boldsymbol{F}_{\mathrm{p}}$ also includes rearrangements of the grains.

The first main assumption of perfect finite plasticity is that the stored–energy density (per unit volume) ψ depends on $\boldsymbol{F}$ and $\boldsymbol{F}_{\mathrm{p}}$ only through the elastic part of the deformation gradient

$$\psi = \widehat{\psi}(\boldsymbol{F}, \boldsymbol{P}) = \Psi(\boldsymbol{F}_{\mathrm{e}}) = \Psi(\boldsymbol{F}\boldsymbol{P}) \text{ where } \boldsymbol{P} = \boldsymbol{F}_{\mathrm{p}}^{-1}. \tag{2.2}$$

For notational simplicity we omit all dependence on the material point $x \in \Omega$, but of course our theory can handle general inhomogeneous materials with $\psi = \Psi(x, \boldsymbol{F}\boldsymbol{P})$.

Using the elastic stress tensor (first Piola–Kirchhoff tensor)

$$\boldsymbol{\Sigma} = \tfrac{\partial}{\partial \boldsymbol{F}}\widehat{\psi}(\boldsymbol{F}, \boldsymbol{P}) = \mathrm{D}\Psi(\boldsymbol{F}\boldsymbol{P})\boldsymbol{P}^\top \tag{2.3}$$

the elastic equilibrium equation reads

$$\begin{cases} -\mathrm{div}\boldsymbol{\Sigma} = \boldsymbol{f}_{\mathrm{ext}}(t, \cdot) \text{ in } \Omega, \\ \boldsymbol{\varphi} \quad = \boldsymbol{\varphi}_0(t, \cdot) \quad \text{on } \Gamma_{\mathrm{displ}}, \\ \boldsymbol{\Sigma}\boldsymbol{\nu} \quad = \boldsymbol{t}_{\mathrm{ext}}(t, \cdot) \quad \text{on } \Gamma_{\mathrm{tract}} = \partial\Omega/\Gamma_{\mathrm{displ}}. \end{cases} \tag{2.4}$$

We additionally need a flow rule for the internal plastic variable $\boldsymbol{P} = \boldsymbol{F}_{\mathrm{p}}^{-1}$. This is obtained from the principle of maximal dissipation or equivalently from the associated flow rule. The aim of this section is to derive another formulation using a dissipation metric and the calculus of subgradients.

We start with the conventional formulation. The plastic back stress $\boldsymbol{Q}$ can be defined as

$$\boldsymbol{Q} = -\tfrac{\partial}{\partial \boldsymbol{P}}\widehat{\psi}(\boldsymbol{F}, \boldsymbol{P}) = -\boldsymbol{F}^\top \mathrm{D}\Psi(\boldsymbol{F}\boldsymbol{P}). \tag{2.5}$$

Note that the tensor $\overline{\boldsymbol{Q}} = \boldsymbol{P}^\top\boldsymbol{Q}$ depends only on the elastic part $\boldsymbol{F}_{\mathrm{e}} = \boldsymbol{F}\boldsymbol{P}$ of the deformations.

The yield function y in general depends on $\boldsymbol{Q}$ and $\boldsymbol{P}$ and $y = 0$ defines the yield surface. Again assuming that y is invariant under all plastic deformations we postulate

$$y = \widehat{y}(\boldsymbol{Q}, \boldsymbol{P}) = Y(\overline{\boldsymbol{Q}}) = Y(\boldsymbol{P}^\top\boldsymbol{Q}). \tag{2.6}$$

The principle of maximal dissipation generates the associated flow rule (cf. [SiO85, ZiW87, Sim88])

$$\boldsymbol{P}^{-1}\dot{\boldsymbol{P}} = \lambda\tfrac{\partial Y}{\partial \overline{\boldsymbol{Q}}}(\overline{\boldsymbol{Q}}) \text{ with } \lambda \geq 0,\ Y(\overline{\boldsymbol{Q}}) \leq 0 \text{ and } \lambda Y(\overline{\boldsymbol{Q}}) = 0. \tag{2.7}$$

Now (2.7) together with the elastic equilibrium condition (2.4) form the full set of equations describing the evolution of elastoplasticity.

The flow rule (2.7), which now has the Karush–Kuhn–Tucker form, can be reformulated by introducing

$$\mathbb{Q} = \{\overline{\boldsymbol{Q}} \in \mathbb{R}^{d\times d} : Y(\overline{\boldsymbol{Q}}) \le 0\} \text{ and } \mathcal{X}_{\mathbb{Q}}(\overline{\boldsymbol{Q}}) = \begin{cases} 0 & \text{for } \overline{\boldsymbol{Q}} \in \mathbb{Q} \\ \infty & \text{else.} \end{cases}$$

where we always assume that

$$\mathbb{Q} \text{ is convex and there exist } r_1, r_2 > 0 \text{ such that } B(0,r_1) \subset \mathbb{Q} \subset B(0,r_2). \tag{2.8}$$

The flow rule (2.7) now takes the form

$$\boldsymbol{P}^{-1}\dot{\boldsymbol{P}} \in \partial\mathcal{X}_{\mathbb{Q}}(\overline{\boldsymbol{Q}}) = \mathrm{N}_{\overline{\boldsymbol{Q}}}\mathbb{Q}. \tag{2.9}$$

Note that (2.9) is more general, since $\mathbb{Q}$ may have corners while $\partial\mathbb{Q} = \{\overline{\boldsymbol{Q}} : Y(\overline{\boldsymbol{Q}}) = 0\}$ is C^1 if $Y \in \mathrm{C}^1$.

For general convex functions $f : Z \to \mathbb{R}$ on a Banach space Z the subdifferential $\partial f(z)$ at the point $z \in Z$ is defined via

$$\partial f(z) = \{z^* \in Z^* : f(z+\widehat{z}) \ge f(z) + \langle z^*, \widehat{z}\rangle \text{ for all } \widetilde{z} \in Z\}, \tag{2.10}$$

where Z^* is the dual Banach space and $\langle z^*, z\rangle$ the dual pairing. The outer normal cone $\mathrm{N}_z C$ of a closed convex set C in the point $z \in C$ is

$$\mathrm{N}_z C = \{z^* \in Z^* : z^*(z-\widehat{z}) \le 0 \text{ for all } \widehat{z} \in C\} \tag{2.11}$$

Using the Legendre transform $(\mathcal{L}f)(z^*) = \sup_{z\in Z}(\langle z^*, z\rangle - f(z))$ we define the dissipation functional

$$\Delta(\boldsymbol{\xi}) = (\mathcal{L}\mathcal{X}_{\mathbb{Q}})(\boldsymbol{\xi}) = \sup\Big\{\langle \boldsymbol{\eta}, \boldsymbol{\xi}\rangle \ : \ \boldsymbol{\eta} \in \mathbb{Q}\Big\}. \tag{2.12}$$

With (2.8) we conclude $r_1\|\boldsymbol{\xi}\| \le \Delta(\boldsymbol{\xi}) \le r_2\|\boldsymbol{\xi}\|$. Moreover $\Delta : \mathbb{R}^{d\times d} \to [0,\infty)$ is convex (as $\mathbb{Q}$ is convex) and homogeneous of degree 1. Under the additional assumption $\Delta(-\boldsymbol{\xi}) = \Delta(\boldsymbol{\xi})$, which is not necessary for the present work, the function Δ defines a metric on $\mathbb{R}^{d\times d}$. Without this condition we call Δ a **nonsymmetric** metric.

We define the dissipation functional

$$\widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}) = \Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}).$$

The classical duality theory for convex functions applied to $\Delta = \mathcal{L}\mathcal{X}_{\mathbb{Q}}$ gives

$$\boldsymbol{\xi} \in \partial\mathcal{X}_{\mathbb{Q}}(\overline{\boldsymbol{Q}}) \quad\Longleftrightarrow\quad \overline{\boldsymbol{Q}} \in \partial\Delta(\boldsymbol{\xi}).$$

With $\boldsymbol{\xi} = \boldsymbol{P}^{-1}\dot{\boldsymbol{P}}$ and $\partial_2\widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}) = \boldsymbol{P}^{-\top}\partial\Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}})$ we find two equivalent forms of the flow rule

$$\overline{\boldsymbol{Q}} \in \partial\Delta(\boldsymbol{P}^{-1}\dot{P}) \quad\Longleftrightarrow\quad \boldsymbol{Q} \in \partial_2\widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}). \tag{2.13}$$

We employ a standard relation for subdifferentials of convex functions f which are homogeneous of degree 1:

$$\partial f(z) = \{\, z^* \in Z^* \,:\, z^* \in \partial f(0) \text{ and } \langle z^*, z\rangle = f(z)\,\}. \tag{2.14}$$

Together with the definition of $\boldsymbol{Q}$ in (2.5) the right–hand side of (2.13) gives our final pointwise version of the flow rule:

$$\mathbf{0} \in \tfrac{\partial}{\partial \boldsymbol{P}}\widehat{\psi}(\boldsymbol{F},\boldsymbol{P}) + \partial_2\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) \subset \mathbb{R}^{d\times d}. \tag{2.15}$$

To obtain field equations we consider the integrated functionals

$$\begin{aligned}&\mathcal{E}(t,\boldsymbol{\varphi},\boldsymbol{P}) = \textstyle\int_\Omega[\widehat{\psi}(\mathrm{D}_x\boldsymbol{\varphi}(x),\boldsymbol{P}(x)) - \boldsymbol{f}_{\mathrm{ext}}(t,x)\boldsymbol{\cdot}\boldsymbol{\varphi}(x)]\,\mathrm{d}x - \int_{\Gamma_{\mathrm{tract}}}\boldsymbol{t}_{\mathrm{ext}}(t,x)\boldsymbol{\cdot}\boldsymbol{\varphi}(x)\,\mathrm{d}a,\\ &\widehat{\Delta}_\Omega(\boldsymbol{P},\dot{\boldsymbol{P}}) = \textstyle\int_\Omega \widehat{\Delta}(\boldsymbol{P}(x),\dot{\boldsymbol{P}}(x))\,\mathrm{d}x.\end{aligned}$$

Now, the whole problem of elastoplasticity can be written using variational derivatives and the subdifferential as follows:

$$\begin{aligned}&\text{elastic equilibrium:} && \mathrm{D}_{\boldsymbol{\varphi}}\mathcal{E}(t,\boldsymbol{\varphi},\boldsymbol{P}) = 0,\\ &\text{plastic flow rule:} && \mathbf{0} \in \mathrm{D}_{\boldsymbol{P}}\mathcal{E}(t,\boldsymbol{\varphi},\boldsymbol{P}) + \partial_2\widehat{\Delta}_\Omega(\boldsymbol{P},\dot{\boldsymbol{P}}).\end{aligned} \tag{2.16}$$

Our final formulation, which we derive now, has the major advantage that it does not involve derivatives but relies only on minimization of functionals.

Applying (2.14) with $z^* = \boldsymbol{Q} = -\frac{\partial}{\partial \boldsymbol{P}}\widehat{\psi}(\boldsymbol{F},\boldsymbol{P})$ and $f = \widehat{\Delta}(P,\cdot)$ we can calculate the time derivative of the bulk energy $e(t) = \mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t))$ along any solution $t \mapsto (\boldsymbol{\varphi}(t),\boldsymbol{P}(t))$:

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}e(t) &= \tfrac{\partial}{\partial t}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) + \mathrm{D}_{\boldsymbol{\varphi}}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t))[\dot{\boldsymbol{\varphi}}(t)] + \mathrm{D}_{\boldsymbol{P}}\mathcal{E}(t,\boldsymbol{\varphi},\boldsymbol{P})[\dot{\boldsymbol{P}}(t)]\\ &= \tfrac{\partial}{\partial t}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) + 0 + \textstyle\int_\Omega -\langle \boldsymbol{Q}(t,x),\dot{\boldsymbol{P}}(t,x)\rangle\,\mathrm{d}x\\ &= \tfrac{\partial}{\partial t}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) - \textstyle\int_\Omega \widehat{\Delta}(\boldsymbol{P}(t,x),\dot{\boldsymbol{P}}(t,x))\,\mathrm{d}x\\ &= \tfrac{\partial}{\partial t}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) - \widehat{\Delta}_\Omega(\boldsymbol{P}(t),\dot{\boldsymbol{P}}(t)).\end{aligned}$$

This leads to the energy balance

$$\begin{aligned}&\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) - \mathcal{E}(s,\boldsymbol{\varphi}(s),\boldsymbol{P}(s)) + \textstyle\int_s^t \widehat{\Delta}_\Omega(\boldsymbol{P}(\tau),\dot{\boldsymbol{P}}(\tau))\,\mathrm{d}\tau\\ &= \textstyle\int_s^t \frac{\partial}{\partial t}\mathcal{E}(\tau,\boldsymbol{\varphi}(\tau),\boldsymbol{P}(\tau))\,\mathrm{d}\tau\\ &= -\textstyle\int_s^t \int_\Omega \frac{\partial}{\partial t}\boldsymbol{f}_{\mathrm{ext}}(\tau,x)\cdot\boldsymbol{\varphi}(\tau,x)\,\mathrm{d}x + \int_{\Gamma_{\mathrm{tract}}}\frac{\partial}{\partial t}\boldsymbol{t}_{\mathrm{ext}}(\tau,x)\cdot\boldsymbol{\varphi}(\tau,x)\,\mathrm{d}a\,\mathrm{d}\tau.\end{aligned} \tag{2.17}$$

Moreover (2.16) implies the local conditions

$$\begin{aligned}&\mathrm{D}_{\boldsymbol{\varphi}}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t))[\boldsymbol{\psi}] = 0\\ &\mathrm{D}_{\boldsymbol{P}}\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t))[\boldsymbol{S}] + \widehat{\Delta}_\Omega(\boldsymbol{P}(t),\boldsymbol{S}) \geq 0\end{aligned} \tag{2.18}$$

for all $\boldsymbol{\psi}$ with $\boldsymbol{\psi}|_{\Gamma_{\mathrm{displ}}} = \boldsymbol{\varphi}_0(t,\cdot)$ and all $\boldsymbol{S}$. Note that (2.17) and (2.18) are equivalent to (2.16) and hence to (2.4) and (2.7).

Since we are considering rate–independent elastoplasticity we may interpret (2.18) as a local form of a stability condition. In fact, we may assume that $\boldsymbol{\varphi}(t)$ is the (global) minimizer of $\mathcal{E}(t,\cdot,\boldsymbol{P}(t))$. Similarly we may assume that $\boldsymbol{P}(t)$ is such that under the given loading (i.e., t fixed) we cannot lower the energy from $\mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t))$ to a value $\mathcal{E}(t,\boldsymbol{\varphi}^*,\boldsymbol{P}^*)$ such that the gain in the energy is larger than the associated dissipation obtained from moving $\boldsymbol{P}(t)$ continuously into $\boldsymbol{P}^*$.

To measure this dissipation we introduce a nonsymmetric distance metric on matrices via

$$\widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_1) = \inf\Big\{ \textstyle\int_0^1 \hat{\Delta}(\boldsymbol{P}(\tau),\dot{\boldsymbol{P}}(\tau))\,\mathrm{d}\tau : \boldsymbol{P}\in \mathrm{C}^1([0,1],\mathbb{R}^{d\times d}), \tag{2.19}$$

$$\boldsymbol{P}(0)=\boldsymbol{P}_0, \boldsymbol{P}(1)=\boldsymbol{P}_1\Big\}. \tag{2.20}$$

Note that we have $\widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_1) = \widehat{D}(I,\boldsymbol{P}_0^{-1}\boldsymbol{P}_1)$ but we do not assume symmetry, i.e., we may have $\widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_1)\neq \widehat{D}(\boldsymbol{P}_1,\boldsymbol{P}_0)$. The triangle inequality

$$\widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_2)\le \widehat{D}(\boldsymbol{P}_0,\boldsymbol{P}_1)+\widehat{D}(\boldsymbol{P}_1,\boldsymbol{P}_2) \tag{2.21}$$

is an immediate consequence of the definition. Moreover we have $\widehat{D}(I,\mathrm{e}^{\boldsymbol{\xi}}) = \Delta(\boldsymbol{\xi}) + O(|\boldsymbol{\xi}|^2)$ for $\boldsymbol{\xi}\to 0$.

Defining the integrated version of the metric $\widehat{D}_\Omega(\boldsymbol{P}_0,\boldsymbol{P}_1) = \int_\Omega \widehat{D}(\boldsymbol{P}_0(x),\boldsymbol{P}_1(x))\,\mathrm{d}x$ we arrive at our final global formulation of elastoplasticity

$$\begin{array}{ll}\textbf{global stability:} & \mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) \le \mathcal{E}(t,\boldsymbol{\varphi}^*,\boldsymbol{P}^*) + \widehat{D}_\Omega(\boldsymbol{P}(t),\boldsymbol{P}^*) \\ & \text{for all } \boldsymbol{\varphi}^* \text{ with } \boldsymbol{\varphi}^*|_{\Gamma_{\text{displ}}} = \boldsymbol{\varphi}_0(t,\cdot) \text{ and all } \boldsymbol{P}^*; \\ \textbf{energy balance:} & \mathcal{E}(t,\boldsymbol{\varphi}(t),\boldsymbol{P}(t)) + \int_s^t \Delta_\Omega(\boldsymbol{P}(\tau)^{-1}\dot{\boldsymbol{P}}(\tau))\,\mathrm{d}\tau \\ & = \mathcal{E}(s,\boldsymbol{\varphi}(s),\boldsymbol{P}(s)) + \int_s^t \frac{\partial}{\partial t}\mathcal{E}(\tau,\boldsymbol{\varphi}(\tau),\boldsymbol{P}(\tau))\,\mathrm{d}\tau. \end{array} \tag{2.22}$$

Clearly this formulation implies the previous ones but not vice versa. Equivalence can be obtained, however, under certain convexity assumption like in linear elastoplasticity.

A major advantage of the formulation (2.22) is that it immediately gives rise to a suitable incremental formulation when the time interval $[0,T]$ is discretized via $0=t_0< t_1<\ldots<t_N=T$ and the initial value $(\boldsymbol{\varphi}_0,\boldsymbol{P}_0)$ at $t_0=0$ is given:

$$\begin{array}{l}\textbf{Incremental Problem}\text{: Given } (\boldsymbol{\varphi}_k,\boldsymbol{P}_k) \text{ at time } t_k, \\ \text{find } (\boldsymbol{\varphi}_{k+1},\boldsymbol{P}_{k+1}) \text{ such that it minimizes the} \\ \text{function } T_{k+1} : (\boldsymbol{\varphi},\boldsymbol{P}) \mapsto \mathcal{E}(t_{k+1},\boldsymbol{\varphi},\boldsymbol{P}) + \widehat{D}_\Omega(\boldsymbol{P}_k,\boldsymbol{P}).\end{array} \tag{2.23}$$

We show that any solutions of (2.23) satisfy a suitable discretized version of (2.22).

Lemma 2.1 *Let $(\boldsymbol{\varphi}_k,\boldsymbol{P}_k)$, $k=1,\ldots,N$, be a solution of the incremental problem (2.23). Then, for $k=1,\ldots,N$ we have*

(i) stability of $(\boldsymbol{\varphi}_k,\boldsymbol{P}_k)$ at time t_k, i.e.,

$$\mathcal{E}(t_k,\boldsymbol{\varphi}_k,\boldsymbol{P}_k)\le \mathcal{E}(t,\boldsymbol{\varphi}^*,\boldsymbol{P}^*) + \widehat{D}_\Omega(\boldsymbol{P}_k,\boldsymbol{P}^*) \textit{ for all } (\boldsymbol{\varphi}^*,\boldsymbol{P}^*),$$

(ii) the two–sided discretized energy estimate

$$\begin{aligned}&\mathcal{E}(t_k,\boldsymbol{\varphi}_k,\boldsymbol{P}_k)-\mathcal{E}(t_{k-1},\boldsymbol{\varphi}_k,\boldsymbol{P}_k)=\textstyle\int_{t_{k-1}}^{t_k}\frac{\partial}{\partial t}\mathcal{E}(s,\boldsymbol{\varphi}_k,\boldsymbol{P}_k)\,\mathrm{d}s\\ &\le \mathcal{E}(t_k,\boldsymbol{\varphi}_k,\boldsymbol{P}_k)-\mathcal{E}(t_{k-1},\boldsymbol{\varphi}_{k-1},\boldsymbol{P}_{k-1})-\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}_k)\\ &\le \textstyle\int_{t_{k-1}}^{t_k}\frac{\partial}{\partial t}\mathcal{E}(s,\boldsymbol{\varphi}_{k-1},\boldsymbol{P}_{k-1})\,\mathrm{d}s=\mathcal{E}(t_k,\boldsymbol{\varphi}_{k-1},\boldsymbol{P}_{k-1})-\mathcal{E}(t_{k-1},\boldsymbol{\varphi}_{k-1},\boldsymbol{P}_{k-1}).\end{aligned}$$

PROOF: To shorten the notation we let $\boldsymbol{z}_k=(\boldsymbol{\varphi}_k,\boldsymbol{P}_k)$. The stability (i) is obtained as follows. For all $\boldsymbol{z}^*$ we have

$$\begin{aligned}&\mathcal{E}(t_k,\boldsymbol{z}^*)+\widehat{D}_\Omega(\boldsymbol{P}_k,\boldsymbol{P}^*)=\mathcal{E}(t_k,\boldsymbol{z}^*)+\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}^*)+\widehat{D}_\Omega(\boldsymbol{P}_k,\boldsymbol{P}^*)-\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}^*)\\ &\ge \mathcal{E}(t_k,\boldsymbol{z}_k)+\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}_k)+\widehat{D}_\Omega(\boldsymbol{P}_k,\boldsymbol{P}^*)-\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}^*)\ge\mathcal{E}(t_k,\boldsymbol{z}_k).\end{aligned}$$

Here, the first inequality uses that $\boldsymbol{z}_k$ is a global minizer (cf. (2.23)) and the second estimate is the triangle inequality (2.21).

The lower estimate in (ii) follows from the stability of $\boldsymbol{z}_{k-1}$ at time t_{k-1}:

$$\mathcal{E}(t_k,\boldsymbol{z}_k)-\mathcal{E}(t_{k-1},\boldsymbol{z}_{k-1})+\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}_k)\ge\mathcal{E}(t_k,\boldsymbol{z}_k)-\mathcal{E}(t_{k-1},\boldsymbol{z}_k).$$

The upper estimate is a consequence of $\boldsymbol{z}_k$ being a minimizer:

$$\mathcal{E}(t_k,\boldsymbol{z}_k)+\widehat{D}_\Omega(\boldsymbol{P}_{k-1},\boldsymbol{P}_k)\le\mathcal{E}(t_k,\boldsymbol{z}_{k-1})=\mathcal{E}(t_{k-1},\boldsymbol{z}_{k-1})+\textstyle\int_{t_{k-1}}^{t_k}\frac{\partial}{\partial t}\mathcal{E}(s,\boldsymbol{z}_{k-1})\,\mathrm{d}s.$$

This proves the result. $\square$

The minimization problem in the incremental problem (2.23) can be solved in two steps, since the functional T_{k+1} has the form

$$T_{k+1}(\boldsymbol{\varphi},\boldsymbol{P})=\textstyle\int_\Omega\Psi((\mathrm{D}_x\boldsymbol{\varphi})\boldsymbol{P})+\widehat{D}(\boldsymbol{P}_k,\boldsymbol{P})\,\mathrm{d}x-\mathcal{L}_{\mathrm{ext}}(t_{k+1},\boldsymbol{\varphi}),$$

where $\mathcal{L}_{\mathrm{ext}}(t,\cdot)$ denotes the external loading. We first minimize pointwise in $x\in\Omega$ with respect to $\boldsymbol{P}$ and obtain the reduced incremental energy density

$$\Psi^{\mathrm{red}}_{\boldsymbol{P}_k}(\boldsymbol{F})=\min\{\,\Psi(\boldsymbol{F}\boldsymbol{P})+\widehat{D}(\boldsymbol{P}_k,\boldsymbol{P}):\boldsymbol{P}\in G\}.$$

To study the existence question for solutions $(\boldsymbol{\varphi}_{k+1},\boldsymbol{P}_{k+1})$ it is now essential to prove certain convexity properties of $\Psi^{\mathrm{red}}_{\boldsymbol{P}_k}(\cdot):\mathbb{R}^{d\times d}\to[0,\infty]$ such as rank–one and quasi–convexity, see [OrR99, CHM00] for first results in this direction. For the present metric $\widehat{D}$ this will be done in subsequent work.

3 Mathematical formulation using Lie groups

In general situations of plasticity with finite elastic and plastic deformations the deformation gradient $\boldsymbol{F}$ is assumed to lie in $\mathrm{GL}(3)=\mathrm{GL}_+(3,\mathbb{R})=\{\boldsymbol{p}\in\mathbb{R}^{3\times3}:\det\boldsymbol{P}>0\}$. The plastic variable $\boldsymbol{P}=\boldsymbol{F}_{\mathrm{p}}^{-1}$ is assumed to lie in a Lie subgroup G of $\mathrm{GL}(3)$. For example, $G=\mathrm{SL}(3)=\{\boldsymbol{P}\in\mathrm{GL}(3):\det\boldsymbol{P}=1\}$ is the case of incompressible plastic deformations. If only one slip system should be taken into account, then we let $G=\{\boldsymbol{I}+\alpha\boldsymbol{n}\otimes\boldsymbol{m}:\alpha\in\mathbb{R}\}$, where $|\boldsymbol{n}|=|\boldsymbol{m}|=1$ and $\boldsymbol{n}\cdot\boldsymbol{m}=0$. Hence, in the following $G\subset\mathrm{GL}(d)$ will denote a general Lie subgroup which contains the plastic variable: $\boldsymbol{P}(t,x) \in G$. By $\boldsymbol{g} = \mathrm{T}_{\boldsymbol{I}}G$ we denote the associated Lie algebra.

Our plasticity model is solely based on two constitutive functions: the elastic energy density $\widehat{\psi}(\boldsymbol{F},\boldsymbol{P})$ and the plastic dissipation potential $\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}})$,

$$\widehat{\psi} : \mathrm{GL}(d) \times G \to \mathbb{R} \quad \text{ and } \quad \widehat{\Delta} : \mathrm{T}G \to [0,\infty).$$

Here again we neglect any dependence on $x \in \Omega$ which could be included easily. As a consequence the elastic stress $\boldsymbol{\Sigma} = \frac{\partial}{\partial \boldsymbol{F}}\widehat{\psi}(\boldsymbol{F},\boldsymbol{P})$ is an element of $\mathrm{T}^*_{\boldsymbol{F}}\mathrm{GL}(d)$ and the plastic backstress $\boldsymbol{Q} = -\frac{\partial}{\partial \boldsymbol{P}}\widehat{\psi}(\boldsymbol{F},\boldsymbol{P})$ is an element of $\mathrm{T}^*_{\boldsymbol{P}}G$. Hence, the flow rule (2.13) has to be considered on T^*G.

To introduce the (symmetry) axiom of our plasticity model we use the left and right translations $L_{\boldsymbol{G}}$ and $R_{\boldsymbol{G}}$ on $\mathrm{GL}(d)$:

$$L_{\boldsymbol{G}} : \begin{cases} G & \to & G, \\ \boldsymbol{P} & \mapsto & \boldsymbol{G}\boldsymbol{P}; \end{cases} \qquad R_{\boldsymbol{G}} : \begin{cases} G & \to & G, \\ \boldsymbol{P} & \mapsto & \boldsymbol{P}\boldsymbol{G}. \end{cases}$$

The derivative $\mathrm{D}L_{\boldsymbol{G}}$ maps the tangent space $\mathrm{T}_{\boldsymbol{P}}G$ bijectively onto $\mathrm{T}_{L_{\boldsymbol{G}}\boldsymbol{P}}G$. The basic axiom of multiplicative plasticity can now be formulated by introducing an action of G on $\mathrm{GL}(d) \times TG$ via

$$\tau : \begin{cases} G \times (\mathrm{GL}(d) \times \mathrm{T}G) & \to & \mathrm{GL}(d) \times \mathrm{T}G, \\ (\boldsymbol{G},(\boldsymbol{F},\boldsymbol{P},\dot{\boldsymbol{P}})) & \mapsto & (R_{\boldsymbol{G}}^{-1}\boldsymbol{F}, L_{\boldsymbol{G}}\boldsymbol{P}, \mathrm{D}L_{\boldsymbol{G}}\dot{\boldsymbol{P}}). \end{cases}$$

For $\tau(\boldsymbol{G},(\boldsymbol{F},\boldsymbol{P},\dot{\boldsymbol{P}}))$ we use the abbreviation

$$\tau_{\boldsymbol{G}}(\boldsymbol{F},\boldsymbol{P},\dot{\boldsymbol{P}}) = (R_{\boldsymbol{G}}^{-1}\boldsymbol{F}, L_{\boldsymbol{G}}\boldsymbol{P}, \mathrm{D}L_{\boldsymbol{G}}\dot{\boldsymbol{P}}) \quad = [\![\, (\boldsymbol{F}\boldsymbol{G}^{-1}, \boldsymbol{G}\boldsymbol{P}, \boldsymbol{G}\dot{\boldsymbol{P}}) \,]\!] \, .$$

In the whole of this section, we will use the abstract Lie group notation but add for clarity the corresponding matrix notation in the double brackets $[\![\quad]\!]$.

In addition to the plastic symmetry there are two standard symmetries which are already present in pure elasticity. The first symmetry is called objectivity or frame indifference and is due to the Euclidian symmetry of the space $\mathbb{R}^d$. The rotations $\boldsymbol{R} \in \mathrm{O}(d)$ act on $\boldsymbol{F}_{\mathrm{e}}$ via left translation. The second symmetry is the material symmetry which is incorporated in a subgroup S of $\mathrm{O}(d)$ and which acts on $\boldsymbol{F}_{\mathrm{e}}$ via right translation.

(A1) **Objectivity (frame indifference):** For all $\boldsymbol{R} \in \mathrm{SO}(d)$ and all $(\boldsymbol{F},\boldsymbol{P}) \in \mathrm{GL}(d)\times G$ we have

$$\widehat{\psi}(L_{\boldsymbol{R}}\boldsymbol{F},\boldsymbol{P}) = \widehat{\psi}(\boldsymbol{F},\boldsymbol{P}) = [\![\, \widehat{\psi}(\boldsymbol{R}\boldsymbol{F},\boldsymbol{P}) \,]\!] \, .$$

(A2) **Invariance under superimposed plastic deformations:**

$$\widehat{\psi} \circ \left(\tau_{\boldsymbol{G}}|_{\mathrm{GL}(d)\times G}\right) = \widehat{\psi} \; \text{ and } \; \widehat{\Delta} \circ \left(\tau_{\boldsymbol{G}}|_{\mathrm{T}G}\right) = \widehat{\Delta} \text{ for all } \boldsymbol{G} \in G;$$

$$[\![\, \widehat{\psi}(\boldsymbol{F}\boldsymbol{G}^{-1},\boldsymbol{G}\boldsymbol{P}) = \widehat{\psi}(\boldsymbol{F},\boldsymbol{P}) \text{ and } \widehat{\Delta}(\boldsymbol{G}\boldsymbol{P},\boldsymbol{G}\dot{\boldsymbol{P}}) = \widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) \,]\!] \, .$$

(A3) **Rate independency and dissipativity:** The function $\widehat{\Delta}(\boldsymbol{P},\cdot) : \mathrm{T}_{\boldsymbol{P}}G \to [0,\infty)$ is convex and homogeneous of degree 1 with $r_1\|\boldsymbol{\xi}\| \le \widehat{\Delta}(\boldsymbol{P}, L_{\boldsymbol{P}}\boldsymbol{\xi}) \le r_2\|\boldsymbol{\xi}\|$.

(A4) **Material symmetry:** For for all $\boldsymbol{S} \in S$ and all $(\boldsymbol{F}, \boldsymbol{P}, \dot{\boldsymbol{P}}) \in \mathrm{GL}(d) \times \mathrm{T}G$ we have

$$\left.\begin{array}{rclcl} \widehat{\psi}(\boldsymbol{F}, R_{\boldsymbol{S}}\boldsymbol{P}) & = & \widehat{\psi}(\boldsymbol{F}, \boldsymbol{P}) & = & [\![\, \widehat{\psi}(\boldsymbol{F}, \boldsymbol{P}\boldsymbol{S}) \,]\!] \, , \\ \widehat{\Delta}(R_{\boldsymbol{S}}\boldsymbol{P}, \mathrm{D}R_{\boldsymbol{S}}\dot{\boldsymbol{P}}) & = & \widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}) & = & [\![\, \widehat{\Delta}(\boldsymbol{P}\boldsymbol{S}, \dot{\boldsymbol{P}}\boldsymbol{S}) \,]\!] \, . \end{array}\right\}$$

Assumptions (A2) and (A3) imply

$$\widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}) = \Delta(\mathrm{D}L_{\boldsymbol{P}}^{-1}\dot{\boldsymbol{P}}) \quad = [\![\, \Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}) \,]\!] \, ,$$

where $\Delta : \boldsymbol{g} = \mathrm{T}_{\boldsymbol{I}}G \to [0, \infty)$ is convex and homogeneous of degree 1. Assumptions (A1) and (A2) give

$$\widehat{\psi}(\boldsymbol{F}, \boldsymbol{P}) = \Psi(\boldsymbol{F}\boldsymbol{P}) = \widetilde{\Psi}((\boldsymbol{F}\boldsymbol{P})^\top \boldsymbol{F}\boldsymbol{P}).$$

The material symmetry $S \subset \mathrm{O}(d)$ in (A4) reads for the reduced functionals Δ and Ψ

$$\Delta(\mathrm{Ad}_{\boldsymbol{S}}\boldsymbol{\xi}) = \Delta(\boldsymbol{\xi}) \text{ and } \Psi(\boldsymbol{F}_{\mathrm{e}}\boldsymbol{S}) = \Psi(\boldsymbol{F}_{\mathrm{e}}) \text{ for } \boldsymbol{S} \in S, \, \boldsymbol{\xi} \in \boldsymbol{g} \text{ and } \boldsymbol{F}_{\mathrm{e}} \in \mathrm{GL}(d). \tag{3.1}$$

Here $\mathrm{Ad}_{\boldsymbol{R}}\boldsymbol{\xi} = \mathrm{D}L_{\boldsymbol{R}}\mathrm{D}R_{\boldsymbol{R}}^{-1}\boldsymbol{\xi} = [\![\, \boldsymbol{R}\boldsymbol{\xi}\boldsymbol{R}^{-1} \,]\!]$ is the adjoint action of G on $\boldsymbol{g}$. We will also need the co–adjoint action of G on $\boldsymbol{g}^*$ which is defined by $\boldsymbol{G} \mapsto \mathrm{CoAd}_{\boldsymbol{G}} = \mathrm{Ad}^*_{\boldsymbol{G}^{-1}} = \mathrm{Ad}^{-*}_{\boldsymbol{G}} \in \mathrm{Lin}(\boldsymbol{g}^*)$.

The construction of suitable energy densities Ψ is well understood. Typically one assumes that the mapping $\boldsymbol{F} \mapsto \Psi(\boldsymbol{F}\boldsymbol{P})$ is such that it satisfies the conditions of pure elasticity, that is $\Psi(\boldsymbol{F}_{\mathrm{e}}) = +\infty$ for $\det \boldsymbol{F}_{\mathrm{e}} \leq 0$ and $\Psi(\boldsymbol{F}_e) \to \infty$ for $\det \boldsymbol{F}_{\mathrm{e}} \searrow 0$ or $\|\boldsymbol{F}_{\mathrm{e}}\| \to \infty$. Moreover, Ψ is assumed to be quasi– or polyconvex. The construction of plastic dissipation densities is discussed in the next section.

4 Dissipation functionals

We discuss two different cases, namely von Mises plasticity (cf. [SiO85, Sim88, MiS92]) and single–crystal plasticity (cf. [OrR99, Gur00]). The former is a model for polycrystals such as metals where the presence of grains averages out the different crystallographic directions; the material is assumed to be isotropic, i.e., $S = \mathrm{SO}(d)$.

4.1 von Mises plasticity

The Lie group G of plastic directors is given by the special linear group

$$G = \mathrm{SL}(d) = \{\boldsymbol{P} \in \mathrm{GL}(d) : \det \boldsymbol{P} = 1\}.$$

The associated Lie algebra $\boldsymbol{g} = \mathrm{T}_{\boldsymbol{I}}G$ is

$$\mathrm{sl}(d) = \{\boldsymbol{\xi} \in \mathbb{R}^{d\times d} : \mathrm{trace}\boldsymbol{\xi} = 0\}.$$

The material symmetry is given by

$$S = \mathrm{SO}(d) = \{\boldsymbol{R} \in \mathrm{GL}(d) : \boldsymbol{R}^\top \boldsymbol{R} = \boldsymbol{I}\}.$$

Denoting by $\|\boldsymbol{\xi}\|_F$ the matrix Frobenius norm

$$\|\boldsymbol{\xi}\|_F^2 = \boldsymbol{\xi} : \boldsymbol{\xi} = \mathrm{trace}(\boldsymbol{\xi}\boldsymbol{\xi}^\top) = \textstyle\sum_{j,k=1}^d \xi_{jk}^2$$

we obtain $\|\mathrm{Ad}_{\boldsymbol{R}}\boldsymbol{\xi}\|_F = \|\boldsymbol{\xi}\|_F$ for $\boldsymbol{R} \in \mathrm{SO}(d)$ and $\boldsymbol{\xi} \in \mathbb{R}^{d\times d}$. Moreover we have the following result.

Proposition 4.1 *Let* $\boldsymbol{\xi}_{\mathrm{sym}} = \frac{1}{2}(\boldsymbol{\xi}+\boldsymbol{\xi}^\top)$ *and* $\boldsymbol{\xi}_{\mathrm{anti}} = \frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\xi}^\top)$ *and take any function* $\delta : \mathbb{R}^2 \to [0,\infty)$ *which is convex and homogeneous of degree 1. Then*

$$\Delta(\boldsymbol{\xi}) = \delta(\|\boldsymbol{\xi}_{\mathrm{sym}}\|_F, \|\boldsymbol{\xi}_{\mathrm{anti}}\|_F)$$

satisfies (3.1) for $S = \mathrm{SO}(d)$.

PROOF: We simply use the fact that $\mathrm{Ad}_{\boldsymbol{R}}$ leaves invariant the splitting $\mathbb{R}^{d\times d} = \mathrm{Sym}(d) \oplus \mathrm{Anti}(d)$. $\square$

We consider the special case that Δ is given by a scalar product, i.e.

$$\Delta_{\mathbb{B}}(\boldsymbol{\xi}) = (\langle \mathbb{B}\boldsymbol{\xi}, \boldsymbol{\xi}\rangle)^{1/2} \geq \gamma\|\boldsymbol{\xi}\|_F \tag{4.1}$$

where $\mathbb{B} : \boldsymbol{g} \to \boldsymbol{g}^*$ is symmetric and positive definite ($\gamma > 0$). This case is equivalent to assuming that G is equipped with a left–invariant Riemannian metric. From standard invariant theory it follows that every symmetric operator $\mathbb{B}$ which is Ad–invariant, i.e. $\Delta_{\mathbb{B}}(\mathrm{Ad}_{\boldsymbol{R}}\boldsymbol{\xi}) = \Delta_{\mathbb{B}}(\boldsymbol{\xi})$ for $\boldsymbol{R} \in \mathrm{SO}(d)$ and $\boldsymbol{\xi} \in \mathbb{R}^{d\times d}$, there exist real number γ_1, γ_2 and γ_3 such that

$$\mathbb{B}\boldsymbol{\xi} = \gamma_1 \mathrm{trace}(\boldsymbol{\xi}) I + \gamma_2 \boldsymbol{\xi}_{\mathrm{sym}} + \gamma_3 \boldsymbol{\xi}_{\mathrm{anti}}. \tag{4.2}$$

Here positive definiteness is equivalent to

$$d\gamma_1 + \gamma_2 > 0,\ \gamma_2 > 0 \text{ and } \gamma_3 > 0.$$

Moreover, for $\boldsymbol{\xi} \in \boldsymbol{g} = \mathrm{sl}(d)$ we have $\mathrm{trace}\boldsymbol{\xi} = 0$ and hence γ_1 is irrelevant. On $\boldsymbol{g} = \mathrm{sl}(d)$ we have

$$\Delta_{\mathbb{B}}(\boldsymbol{\xi}) = (\gamma_2\|\boldsymbol{\xi}_{\mathrm{sym}}\|_F^2 + \gamma_3\|\boldsymbol{\xi}_{\mathrm{anti}}\|_F^2)^{1/2}. \tag{4.3}$$

Note that the plastic flow rule gives $\dot{\boldsymbol{P}} = \boldsymbol{P}(\boldsymbol{\xi}_{\mathrm{sym}} + \boldsymbol{\xi}_{\mathrm{anti}})$ such that $\boldsymbol{\xi}_{\mathrm{sym}}$ corresponds to plastic stretching whereas $\boldsymbol{\xi}_{\mathrm{anti}}$ corresponds to plastic spin. The controversy about the relevance of plastic spin can be rephrased in our setting in the relative size of γ_3 with respect to γ_2. We will discuss this question after the construction of the geodesic curves.

4.2 Single–cristal plasticity

In single crystals plasticity occurs through movements of dislocation and the plastic strain $\boldsymbol{F}_{\mathrm{p}} = \boldsymbol{P}^{-1}$ can be thought of as atomic disarrangements. The motion of dislocations is viewed to take place on n crystallographically determined slip systems $\boldsymbol{S}^\alpha = \boldsymbol{d}^\alpha \otimes \boldsymbol{n}^\alpha$, $\alpha = 1, \ldots, n$, where $\boldsymbol{n}^\alpha$ is the unit normal vector to the αth slip plane and $\boldsymbol{d}^\alpha$ is the associated slip direction with $|\boldsymbol{d}^\alpha| = 1$ and $\boldsymbol{d}^\alpha \cdot \boldsymbol{n}^\alpha = 0$. The plastic flow then takes the form (see [Gur00])

$$\dot{\boldsymbol{F}}_{\mathrm{p}} = (\textstyle\sum_{\alpha=1}^n \nu_\alpha \boldsymbol{S}^\alpha)\boldsymbol{F}_{\mathrm{p}}$$

where $\nu_\alpha \geq 0$ is the slip rate of the slip system $\boldsymbol{S}^\alpha$ (negative slip rates are realized here by using the negative slip system $-\boldsymbol{S}^\alpha$). The crystal symmetry group $S \subset \mathrm{O}(d)$ associates a permutation $\pi_{\boldsymbol{R}} \in \mathrm{Perm}(n)$ to each $\boldsymbol{R} \in S$ such that $\pi_{\boldsymbol{R}\circ\widehat{\boldsymbol{R}}} = \pi_{\boldsymbol{R}} \circ \pi_{\widehat{\boldsymbol{R}}}$ and

$$(\boldsymbol{R}\boldsymbol{d}^\alpha, \boldsymbol{R}\boldsymbol{n}^\alpha) = (\boldsymbol{d}^{\pi_{\boldsymbol{R}}(\alpha)}, \boldsymbol{n}^{\pi_{\boldsymbol{R}}(\alpha)}) \quad \Longleftrightarrow \quad \boldsymbol{S}^{\pi_{\boldsymbol{R}}(\alpha)} = \boldsymbol{R}\boldsymbol{S}^\alpha\boldsymbol{R}^\top \text{ for } \alpha = 1, \ldots, n.$$

For our plastic director $\boldsymbol{P} = \boldsymbol{F}_\mathrm{p}^{-1}$ we obtain $\dot{\boldsymbol{P}} = -\boldsymbol{P}\sum_{\alpha=1}^n \nu_\alpha \boldsymbol{S}^\alpha$, which shows that

$$\mathrm{span}\{\boldsymbol{S}^\alpha : \alpha = 1, \dots, n\} \subset \boldsymbol{g} = \mathrm{T}_{\boldsymbol{I}} G.$$

This generates in a natural way the associated Lie group via the smallest Lie algebra $\boldsymbol{g}$ containing all the slip systems. From $\boldsymbol{d}^\alpha \cdot \boldsymbol{n}^\alpha = 0$ we know $\mathrm{trace}(\boldsymbol{S}^\alpha) = 0$ and hence $\boldsymbol{g} \subset \mathrm{sl}(d)$ and $G \subset \mathrm{SL}(d)$.

According to [Gur00], Section 4.2, the dissipation associated to $\boldsymbol{\xi} = -\sum_{\alpha=1}^n \nu_\alpha \boldsymbol{S}^\alpha$ is given by $\sum_{\alpha=1}^n \kappa_\alpha |\nu_\alpha|$ with positive constants κ_α satisfying the material symmetry relation $\kappa_{\pi_{\boldsymbol{R}}(\alpha)} = \kappa_\alpha$ for all $\boldsymbol{R} \in S$. Since in general the coefficients ν_α are not uniquely determined by $\boldsymbol{\xi}$ we define

$$\Delta(\boldsymbol{\xi}) = \min\{\, \textstyle\sum_{\alpha=1}^n \kappa_\alpha |\nu_\alpha| : \boldsymbol{\xi} = -\sum_{\alpha=1}^n \nu_\alpha \boldsymbol{S}^\alpha \,\}.$$

To match this construction with our abstract theory we need to show $r_1\|\boldsymbol{\xi}\| \le \Delta(\boldsymbol{\xi}) \le r_2\|\boldsymbol{\xi}\|$ for all $\boldsymbol{\xi} \in \boldsymbol{g}$. The lower estimate is trivial with $r = \min_\alpha \kappa_\alpha$, since $\|\boldsymbol{\xi}\|_F \le \sum_{\alpha=1}^n |\nu_\alpha| \|\boldsymbol{S}^\alpha\|_F = \sum_{\alpha=1}^n |\nu_\alpha| \le (1/r)\Delta(\boldsymbol{\xi})$. The upper estimate holds if and only if $\boldsymbol{g}$ equals the span of all $\boldsymbol{S}^\alpha$.

The latter condition is nontrivial as is seen for the case $d = 3$ and $n = 2$ with $\boldsymbol{S}^1 = \boldsymbol{e}_1 \otimes \boldsymbol{e}_2$ and $\boldsymbol{S}^2 = \boldsymbol{e}_2 \otimes \boldsymbol{e}_3$. Clearly, $\mathrm{span}\{\boldsymbol{S}^1, \boldsymbol{S}^2\}$ is two–dimensional but the smallest Lie algebra $\boldsymbol{g}$ containing this span is $\mathrm{span}\{\boldsymbol{S}^1, \boldsymbol{S}^2, \boldsymbol{S}^3\}$ with $\boldsymbol{S}^3 = [\boldsymbol{S}^1, \boldsymbol{S}^2] = \boldsymbol{S}^1\boldsymbol{S}^2 - \boldsymbol{S}^2\boldsymbol{S}^1 = \boldsymbol{e}_1 \otimes \boldsymbol{e}_3$.

5 The metric and geodesics on G

To formulate the elastoplastic problem in Section 2 we used the distance metric $\widehat{D}$ on G which is induced by the (nonsymmetric) norm Δ on $\boldsymbol{g}$ via the left–invariant Finsler–Minkowski metric $\widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}) = \Delta(\mathrm{D}L_{\boldsymbol{P}}^{-1}\dot{\boldsymbol{P}}) = \Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}})$:

$$\widehat{D}_\Delta(\boldsymbol{P}_0, \boldsymbol{P}_1) = \inf\{I_\Delta(\boldsymbol{P}(\cdot)) : \boldsymbol{P} \in \mathrm{C}^1([0,1], G), \boldsymbol{P}(0) = \boldsymbol{P}_0, \boldsymbol{P}(1) = \boldsymbol{P}_1\}, \tag{5.1}$$

with $I_\Delta(\boldsymbol{P}(\cdot)) = \int_0^1 \Delta(\mathrm{D}L_{\boldsymbol{P}(t)}^{-1}\dot{\boldsymbol{P}}(t))\,\mathrm{d}t$.

As a short notation we introduce

$$D_\Delta : \begin{cases} G & \to \quad [0,\infty) \\ \boldsymbol{P} & \mapsto \quad \widehat{D}_\Delta(\boldsymbol{I}, \boldsymbol{P}) \end{cases} \tag{5.2}$$

such that by left–invariance $\widehat{D}_\Delta(\boldsymbol{P}_0, \boldsymbol{P}_1) = D_\Delta(\boldsymbol{P}_0^{-1}\boldsymbol{P}_1)$. If $\Delta(-\boldsymbol{\xi}) = \Delta(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \boldsymbol{g}$ then $D_\Delta(\boldsymbol{P}^{-1}) = D_\Delta(\boldsymbol{P})$.

In general we cannot expect that there is a C^1 function such that the infimum is attained. We also have to take into account curves with corners, namely $\boldsymbol{P} \in \mathrm{W}^{1,\infty}([0,1], G) = \mathrm{C}^{\mathrm{Lip}}([0,1], G)$ where $\dot{\boldsymbol{P}} \in \mathrm{L}^\infty((0,1))$.

A curve $\boldsymbol{P} \in \mathrm{C}^{\mathrm{Lip}}([0,1], G)$ is called a **shortest path** from $\boldsymbol{P}(0)$ to $\boldsymbol{P}(1)$ (with respect to Δ) if $I_\Delta(\boldsymbol{P}) = \widehat{D}_\Delta(\boldsymbol{P}(0), \boldsymbol{P}(1))$. A curve $\boldsymbol{P} \in \mathrm{C}^{\mathrm{Lip}}((t_1, t_2), G)$ is called a **geodesic curve** (with respect to Δ), if for each $t \in (t_1, t_2)$ there exists $\varepsilon > 0$ such that $\boldsymbol{P}|_{[t-\varepsilon, t+\varepsilon]}$ is a shortest path.

Every part $\boldsymbol{P}|_{[\tau_1, \tau_2]}$ of a shortest path is again a shortest path. Moreover, shortest paths are in general not unique.

Example 5.1 Consider the Lie group $G = \{\boldsymbol{P} = \operatorname{diag}(a,b) \in \mathrm{GL}(2) : a, b > 0\}$ and the left–invariant metric $\widehat{\Delta}(\boldsymbol{P}, \dot{\boldsymbol{P}}) = |\dot{a}|/a + |\dot{b}|/b$.

Then, $D_\Delta(\operatorname{diag}(a,b)) = |\log a|+|\log b|$ and any curve $(t \mapsto \operatorname{diag}(\widehat{a}(t), \widehat{b}(t)))$ is a shortest path if $\widehat{a} : [0,1] \to \mathbb{R}$ and $\widehat{b} : [0,1] \to \mathbb{R}$ are monotone.

A curve is a geodesic if and only if $\widehat{a}$ and $\widehat{b}$ are locally monotone. In particular, nonmonotone $\widehat{a}$ or $\widehat{b}$ is admissible if different monotonicity regions are separated by intervals of constant values. E.g. $(\widehat{a}(t), \widehat{b}(t)) = (\max\{1, 2-|t|\}, 1)$ is a geodesic over $t \in \mathbb{R}$.

The next example, which has been mentioned to the author by D. Mittenhuber [Mit00b], shows that shortest paths for left–invariant Finsler metrics in general must have corners. Here the notion of a corner means that the function $\boldsymbol{P} : [0,1] \to G$ is not C^1 when the parametrization is such that $\Delta(\,\mathrm{D}L^{-1}_{\boldsymbol{P}(t)} \dot{\boldsymbol{P}}(t)) \geq c > 0$.

The existence of nonsmooth shortest paths might indicate that certain experimental fact concerning the switching between different slip systems can be modelled easily using Finsler metrics. This topic will be pursued in future research.

Example 5.2 Consider $G = \left\{ \left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{GL}(2) : a > 0,\, b \in \mathbb{R} \right\}$ with the associated Lie algebra $\boldsymbol{g} = \left\{ \left(\begin{smallmatrix} \alpha & \beta \\ 0 & 0 \end{smallmatrix}\right) \in \mathrm{gl}(2) : \alpha, \beta \in \mathbb{R} \right\}$ and the left–invariant Finsler metric induced by $\Delta\left(\left(\begin{smallmatrix} \alpha & \beta \\ 0 & 0 \end{smallmatrix}\right)\right) = |\alpha|+|\beta|$.

Writing $\boldsymbol{P} = \left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right)$ the Finsler metric reads $\Delta(L^{-1}_{\boldsymbol{P}} \dot{\boldsymbol{P}}) = (|\dot{a}|+|\dot{b}|)/a$. To find the shortest path between $\boldsymbol{P}_0 = \boldsymbol{I}$ and $\boldsymbol{P}_1 = \left(\begin{smallmatrix} a_1 & b_1 \\ 0 & 1 \end{smallmatrix}\right)$ we have to minimize

$$\int_0^1 \left(\tfrac{|\dot{a}(t)|}{a(t)} + \tfrac{|\dot{b}(t)|}{a(t)} \right) \mathrm{d}t \quad \text{subject to } \left(\begin{smallmatrix} a(0) \\ b(0) \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \text{ and } \left(\begin{smallmatrix} a(1) \\ b(1) \end{smallmatrix}\right) = \left(\begin{smallmatrix} a_1 \\ b_1 \end{smallmatrix}\right).$$

The integral over the first term gives the total variation of $\log a(\cdot)$ and is bounded from below by $\log a_{\max} + \log \frac{a_{\max}}{a_1}$ where $a_{\max} = \max\{a(t) : t \in [0,1]\}$. The integral over the second term is bounded from below by $|b(1)-b(0)|/a_{\max} = |b_1|/a_{\max}$.

In fact, these lower bounds are attained by the path

$$(a(t), b(t)) = \begin{cases} (1+3t(a_{\max}-1), 0) & \text{for } t \in [0, 1/3], \\ (a_{\max}, (3t-1)b_1) & \text{for } t \in [1/3, 2/3], \\ (a_{\max}-(a_1-a_{\max})(3t-2), b_1) & \text{for } t \in [2/3, 1]. \end{cases}$$

The shortest path is now obtained by minimizing with respect to $a_{\max} \geq \max\{1, a_1\}$. This gives

$$\begin{aligned} D_\Delta\left(\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right)\right) &= \min\left\{ \log \tfrac{m^2}{a} + \tfrac{|b|}{m} : m \geq \max\{1, a\} \right\} \\ &= \begin{cases} \log \frac{1}{a}+|b| & \text{for } a \in (0,1] \text{ and } |b| \leq 2, \\ \log a+|b|/a & \text{for } a \geq \max\{1, |b|/2\}, \\ \log(b^2/4a)+2 & \text{for } |b| \geq 2 \text{ and } a \leq |b|/2. \end{cases} \end{aligned}$$

In particular, we see that the shortest path is unique (up to reparametrization). Hence, almost all shortest paths have one or two corners. (Note that $\dot{a}\dot{b} \equiv 0$ along all shortest paths.)

For the general case we are able to estimate the dissipation metric from above via

$$D_\triangle(\boldsymbol{P}) \le \sum_{i=1}^{n} \Delta(\boldsymbol{\xi}_i) \text{ if } \boldsymbol{P} = \mathrm{e}^{\boldsymbol{\xi}_1} \dots \mathrm{e}^{\boldsymbol{\xi}_n}.$$

(Let $\boldsymbol{P}(t) = \mathrm{e}^{\boldsymbol{\xi}_1} \dots \mathrm{e}^{\boldsymbol{\xi}_k} \mathrm{e}^{(nt-k)\boldsymbol{\xi}_{k+1}}$ for $t \in [k/n, (k+1)/n]$ and use $\boldsymbol{P}(t)^{-1}\dot{\boldsymbol{P}}(t) = n\boldsymbol{\xi}_{k+1}$.)

In fact, we may approximate $D_\triangle(\boldsymbol{I}, \boldsymbol{P})$ as good as we like by increasing n and optimizing $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$. The simple estimate $D_\triangle(\mathrm{e}^{\boldsymbol{\xi}}) \le \Delta(\boldsymbol{\xi})$ is of restricted use, since the (Lie algebraic) exponential map $\boldsymbol{\xi} \mapsto \mathrm{e}^{\boldsymbol{\xi}}$ from $\boldsymbol{g}$ into G is not surjective in general. For a surjective geodesic exponential map $M_\delta : \boldsymbol{g} \to G$, see Corollary 6.2. Of course we have the local expansion $D_\triangle(\mathrm{e}^{\boldsymbol{\xi}}) = \Delta(\boldsymbol{\xi}) + O(\|\boldsymbol{\xi}\|^2)$ for $\boldsymbol{\xi} \to 0$.

Next we show that there always exists a shortest path between any two points on G. After that we use the left–invariance and a nonsmooth version of Noether's theorem to characterize shortest paths.

Theorem 5.3 *For any $\boldsymbol{P}_1 \in G$ there exists a shortest path $\boldsymbol{P}(\cdot) \in \mathrm{C}^{\mathrm{Lip}}([0,1], G)$ from $\boldsymbol{I}$ to $\boldsymbol{P}_1$.*

Proof: We need to find $\boldsymbol{P} \in \mathrm{C}^{\mathrm{Lip}}([0,1], G)$ with $\boldsymbol{P}(0) = \boldsymbol{I}$, $\boldsymbol{P}(1) = \boldsymbol{P}_1$ and

$$I_\triangle(\boldsymbol{P}(\cdot)) = \inf\{I_\triangle(\boldsymbol{Q}) : \boldsymbol{Q} \in \mathrm{C}^1([0,1], G),\ \boldsymbol{Q}(0) = \boldsymbol{I},\ \boldsymbol{Q}(1) = \boldsymbol{P}_1\}. \tag{5.3}$$

This just means that the infimum is in fact a minimum if the set of candidates is increased from C^1 to $\mathrm{C}^{\mathrm{Lip}}$. The right–hand side in (5.3) is by definition $D_\triangle(\boldsymbol{P}_1)$ and we can choose an infimizing sequence $\boldsymbol{Q}^{(k)}$ in $\mathrm{C}^1([0,1], G)$ with $\boldsymbol{Q}^{(k)}(0) = \boldsymbol{I}$, $\boldsymbol{Q}^{(k)}(1) = \boldsymbol{P}_1$ and $\delta_k := I_\triangle(\boldsymbol{Q}^{(k)}) \to D_\triangle(\boldsymbol{P}_1)$. We now use the rate–independency of the integral and reparametrise the time variable. Let $\tau^{(k)}(t) = c_k \int_0^t [1 + \widehat{\Delta}(\boldsymbol{Q}^{(k)}(s), \dot{\boldsymbol{Q}}^{(k)}(s))]\,\mathrm{d}s$ with $c_k = 1/(1+\delta_k)$, then $\tau^{(k)} \in \mathrm{C}^1([0,1], \mathbb{R})$ satisfies $\tau^{(k)}(0) = 0$, $\tau^{(k)}(1) = 1$ and $\dot{\tau}^{(k)}(t) \ge c_k > 0$. Hence, the inverse function $t = T^{(k)}(\tau)$ exists and we can set $\boldsymbol{P}^{(k)}(\tau) = \boldsymbol{Q}^{(k)}(T^{(k)}(\tau))$.

Clearly, $\boldsymbol{P}^{(k)} \in \mathrm{C}^1([0,1], G)$ with $\boldsymbol{P}^{(k)}(0) = \boldsymbol{I}$, $\boldsymbol{P}^{(k)}(1) = \boldsymbol{P}_1$ and $I_\triangle(\boldsymbol{P}^{(k)}) = \delta_k$ by rate independency. Thus, $(\boldsymbol{P}^{(k)})_{k\in\mathbb{N}}$ is also an infimizing sequence. Additionally we know

$$\widehat{\Delta}(\boldsymbol{P}^{(k)}(\tau), \dot{\boldsymbol{P}}^{(k)}(\tau)) = \dot{T}^{(k)}(\tau) a^{(k)}(\tau) = \tfrac{a^{(k)}(\tau)}{c_k(1+a^{(k)}(\tau))} \le 1 + \delta_k$$

where $a^{(k)}(\tau) = \widehat{\Delta}(\boldsymbol{Q}^{(k)}(T^{(k)}(\tau)), \dot{\boldsymbol{Q}}^{(k)}(T^{(k)}(\tau)))$.

This new infimizing sequence has the advantage that

$$\boldsymbol{\xi}^{(k)} := (t \mapsto (\mathrm{D}L^{-1}_{\boldsymbol{P}^{(k)}(t)} \dot{\boldsymbol{P}}(t))$$

is bounded in the linear space $\mathrm{L}^\infty((0,1), \boldsymbol{g})$. From

$$\boldsymbol{P}^{(k)}(t) = \boldsymbol{I} + \textstyle\int_0^t \boldsymbol{P}^{(k)}(s) \boldsymbol{\xi}^{(k)}(s)\,\mathrm{d}s \tag{5.4}$$

we conclude that $\boldsymbol{P}^{(k)}$ is bounded in the linear space $\mathrm{C}^1([0,1], \mathbb{R}^{d\times d})$.

Thus, we may extract a subsequence $(k_l)_{l\in\mathbb{N}}$, such that $(\boldsymbol{P}^{(k_l)})_{l\in\mathbb{N}}$ converges uniformly on $[0,1]$ to a limit $\boldsymbol{P} \in \mathrm{C}^{\mathrm{Lip}}([0,1], \mathbb{R}^{d\times d})$ and $(\boldsymbol{\xi}^{k_l)})_{l\in\mathbb{N}}$ converges weak* to a limit $\boldsymbol{\xi} \in \mathrm{L}^\infty((0,1), \boldsymbol{g})$ with $\|\boldsymbol{\xi}\|_\infty \le 1 + D_\triangle(\boldsymbol{P}_1)$ (or weakly in $\mathrm{L}^2((0,1), \boldsymbol{g})$). Clearly, the limits $\boldsymbol{P}$ and $\boldsymbol{\xi}$ satisfy (5.4) as well and thus $\boldsymbol{P} \in \mathrm{W}^{1,\infty}([0,1], G)$.

From standard theorems in the calculus of variations (cf. [Bal77], [Dac89]) we know that convexity in $\dot{\boldsymbol{P}}$ implies weak lower semicontinuity and we conclude

$$I_\Delta(\boldsymbol{P}) \leq \liminf_{l\to\infty} I_\Delta(\boldsymbol{P}^{(k_l)}) = \liminf_{l\to\infty} \delta_{k_l} = D_\Delta(P_1).$$

This proves the result. $\square$

We now want to characterize shortest paths by deriving suitable differential equations like those for geodesic curves in Riemannian geometry. To motivate the analysis we first assume that $\Delta : \boldsymbol{g} \to [0,\infty)$ is a C^2–function (and no longer homogeneous of degree 1).

The Euler–Lagrange equation associated to the functional $I_\Delta(\boldsymbol{P})$ is given by

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\Big[(\mathrm{D}_{\dot{\boldsymbol{P}}}\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}})\Big] - \mathrm{D}_{\boldsymbol{P}}\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) = 0. \tag{5.5}$$

In matrix notation we have $\mathrm{D}_{\boldsymbol{P}}\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) = -\boldsymbol{P}^{-\top}\,\mathrm{D}\Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}})\dot{\boldsymbol{P}}^\top\boldsymbol{P}^{-\top} \in \mathrm{T}^*_{\boldsymbol{P}}G$ and $\mathrm{D}_{\dot{\boldsymbol{P}}}\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}}) = \boldsymbol{P}^{-\top}\mathrm{D}\Delta(\boldsymbol{P}^{-1}\dot{\boldsymbol{P}})$. Thus, it is possible to write out the equation explicitly. To obtain a proper formulation in Lie group notation it is advantageous to use the Hamiltonian form on T^*G instead of the Lagrangian on $\mathrm{T}G$. Define the conjugate variables $(\boldsymbol{P},\boldsymbol{Q}) \in \mathrm{T}^*G$ via $\boldsymbol{Q} = \mathrm{D}_{\dot{\boldsymbol{P}}}\widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}})$ and the Hamiltonian $\widehat{H}(\boldsymbol{P},\boldsymbol{Q}) = \langle \boldsymbol{Q},\dot{\boldsymbol{P}}\rangle - \widehat{\Delta}(\boldsymbol{P},\dot{\boldsymbol{P}})$ where $\dot{\boldsymbol{P}}$ has to be eliminated. In our special case denote by $f : \boldsymbol{g}^* \to \boldsymbol{g}$ the inverse of the function $\mathrm{D}\Delta : \boldsymbol{g} \to \boldsymbol{g}^*$ and obtain

$$\begin{aligned}\dot{\boldsymbol{P}} &= \mathrm{D}L_{\boldsymbol{P}}f(\mathrm{D}L_{\boldsymbol{P}}^{-*}\boldsymbol{Q}), \text{ where } \mathrm{D}L_{\boldsymbol{P}}^{-*} = (\mathrm{D}L_{\boldsymbol{P}}^{-1})^*,\\ \widehat{H}(\boldsymbol{P},\boldsymbol{Q}) &= H(\mathrm{D}L_{\boldsymbol{P}}^{-*}\boldsymbol{Q}) \text{ with } H(\boldsymbol{\eta}) = \sup\{\langle\boldsymbol{\eta},\boldsymbol{\xi}\rangle - \Delta(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \boldsymbol{g}\}.\end{aligned}$$

We have $f(\boldsymbol{\eta}) = \mathrm{D}H(\boldsymbol{\eta})$ and the Hamiltonian equations which are equivalent to (5.5) are, in matrix notation,

$$\dot{\boldsymbol{P}} = \boldsymbol{P}\,\mathrm{D}H(\boldsymbol{P}^\top\boldsymbol{Q}), \quad \dot{\boldsymbol{Q}} = -\boldsymbol{Q}[\mathrm{D}H(\boldsymbol{P}^\top\boldsymbol{Q})]^\top. \tag{5.6}$$

To obtain a simple Lie algebraic formulation of the left–invariant system we introduce left–invariant coordinates $(\boldsymbol{P},\boldsymbol{\eta}) \in \mathrm{T}^0G := G \times \boldsymbol{g}^*$ via $\boldsymbol{\eta} = \mathrm{D}L_{\boldsymbol{P}}^{-*}\boldsymbol{Q}$. Moreover, we need the Lie–Poisson structure J_{LP} on $\boldsymbol{g}^*$ with $J_{\mathrm{LP}}(\boldsymbol{\eta}) \in \mathrm{Skew}(\boldsymbol{g},\boldsymbol{g}^*)$ and

$$\langle J_{\mathrm{LP}}(\boldsymbol{\eta})\boldsymbol{\xi}_1,\boldsymbol{\xi}_2\rangle = \langle\boldsymbol{\eta},[\boldsymbol{\xi}_1,\boldsymbol{\xi}_2]\rangle \text{ for all } \boldsymbol{\xi}_1,\boldsymbol{\xi}_2 \in \boldsymbol{g},$$

where $[\boldsymbol{\xi}_1,\boldsymbol{\xi}_2]$ denotes the Lie bracket on $\boldsymbol{g}$. According to [AbM78] (5.5) is equivalent to

$$\dot{\boldsymbol{P}} = \mathrm{D}L_{\boldsymbol{P}}\,\mathrm{D}H(\boldsymbol{\eta}), \qquad \dot{\boldsymbol{\eta}} = J_{\mathrm{LP}}(\boldsymbol{\eta})\,\mathrm{D}H(\boldsymbol{\eta}). \tag{5.7}$$

Clearly, the Hamiltonian $\widehat{H}$ is constant along solutions, namely $\frac{\mathrm{d}}{\mathrm{d}t}H(\mathrm{D}L_{\boldsymbol{P}}^{-*}\boldsymbol{Q}(t)) \equiv 0$. Moreover, the Hamiltonian system is invariant under the G–action $(\widetilde{\tau}_{\boldsymbol{G}})_{\boldsymbol{G}\in G}$ with $\widetilde{\tau}_{\boldsymbol{G}} : (\boldsymbol{P},\boldsymbol{Q}) \mapsto (L_{\boldsymbol{G}}\boldsymbol{P},\, \mathrm{D}L_{\boldsymbol{G}}^{-*}\boldsymbol{Q})$. By Noether's theorem (cf. [AbM78, Mie91]) it follows that the co–adjoint action generates the associated first integrals. Define the momentum map

$$\mathcal{J} : \begin{cases} \mathrm{T}^*G & \to \quad \boldsymbol{g}^*,\\ (\boldsymbol{P},\boldsymbol{Q}) & \mapsto \quad \mathrm{Ad}^*_{\boldsymbol{P}^{-1}}(\mathrm{D}L^*_{\boldsymbol{P}}\boldsymbol{Q}) \;=\; [\![\, \boldsymbol{Q}\boldsymbol{P}^\top \,]\!]\,, \end{cases}$$

then for all solutions we have $\mathcal{J}(\boldsymbol{P}(t),\boldsymbol{Q}(t)) \equiv$ const., see [AbM78]. Here we used the coadjoint action Ad^* of G on g^* which is defined via

$$\langle \mathrm{Ad}^*_{\boldsymbol{G}}\boldsymbol{\eta},\boldsymbol{\xi}\rangle = \langle \boldsymbol{\eta},\mathrm{Ad}_{\boldsymbol{G}}\boldsymbol{\xi}\rangle = \langle \boldsymbol{\eta},\, \mathrm{D}L_{\boldsymbol{G}}\,\mathrm{D}R^{-1}_{\boldsymbol{G}}\boldsymbol{\xi}\rangle \text{ for all } \boldsymbol{\xi},\boldsymbol{\eta}.$$

Note that in matrix notation we have

$$[\![\ \mathrm{Ad}^*_{\boldsymbol{G}}\boldsymbol{\eta} = \mathbb{P}(\boldsymbol{G}^\top\boldsymbol{\eta}\boldsymbol{G}^{-\top})\ ]\!]$$

where $\mathbb{P} : \mathbb{R}^{d\times d} \to \boldsymbol{g}^*$ is the projection defined via

$$\mathrm{trace}(\mathbb{P}\boldsymbol{\eta}\boldsymbol{\xi}^\top) = \mathrm{trace}(\boldsymbol{\eta}\boldsymbol{\xi}^\top) \text{ for all } \boldsymbol{\xi}\in\boldsymbol{g}.$$

Letting $\mathcal{J}(\boldsymbol{P}(t),\boldsymbol{Q}(t)) = \boldsymbol{\eta}\in\boldsymbol{g}^*$ and reintroducing $\dot{\boldsymbol{P}}$ by elimination of $\boldsymbol{Q}$ we find the "half–integrated" form of the Euler–Lagrange equation for left–invariant densities $\widehat{\Delta}$:

$$\mathrm{D}\Delta(\mathrm{D}L^{-1}_{\boldsymbol{P}}\dot{\boldsymbol{P}}) = \mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta} \quad \text{ with a constant } \boldsymbol{\eta}\in\boldsymbol{g}^*. \tag{5.8}$$

In the Hamiltonian context this can be rewritten as

$$\dot{\boldsymbol{P}} = \mathrm{D}L_{\boldsymbol{P}}\,\mathrm{D}H(\mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta}) \quad \text{ with a constant } \boldsymbol{\eta}\in\boldsymbol{g}^*. \tag{5.9}$$

The generalization of the ODEs for the stationary curves of the functional $I_\Delta(P)$ to the case of nonsmooth Δ involves the subgradient $\partial\Delta(\boldsymbol{\xi})$ as defined in (2.10). Recall $\mathbb{Q} = \partial\Delta(0) \subset \boldsymbol{g}^*$ and $\partial\Delta(\boldsymbol{\xi}) = \{\boldsymbol{\eta}\in\mathbb{Q} : \Delta(\boldsymbol{\xi}) = \langle\boldsymbol{\eta},\boldsymbol{\xi}\rangle\}$ from (2.14).

We derive the integrated forms (5.8) and (5.9) for shortest paths by directly adapting the proof of Noether's theorem to subgradient techniques.

Theorem 5.4 *Let $\boldsymbol{P} : [0,1] \to G$ be a shortest path between $\boldsymbol{P}(0)$ and $\boldsymbol{P}(1)$ with $\boldsymbol{P}(0) \neq \boldsymbol{P}(1)$. Then there exists a nonzero $\boldsymbol{\eta}\in\boldsymbol{g}^*$ such that*

$$\mathrm{Ad}^*_{\boldsymbol{P}(t)}\boldsymbol{\eta} \in \partial\Delta(\mathrm{D}L^{-1}_{\boldsymbol{P}(t)}\dot{\boldsymbol{P}}(t)) \text{ for a.e. } t\in[0,1]. \tag{5.10}$$

Equivalently, we have $\mathrm{D}L^{-1}_{\boldsymbol{P}}\dot{\boldsymbol{P}} \in \partial\mathcal{X}_{\mathbb{Q}}(\mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta}) = \mathrm{N}_{\mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta}}\mathbb{Q}$ where $\mathrm{N}_{\widetilde{\boldsymbol{\eta}}}\mathbb{Q}$ denotes the normal cone, cf. (2.11).

Proof: We choose $\boldsymbol{\xi}\in\boldsymbol{g}$ and $\rho\in\mathrm{C}^\infty_\mathrm{c}((0,1),\mathbb{R})$ and define $\boldsymbol{Q}_\rho(t) = \mathrm{e}^{\rho(t)\boldsymbol{\xi}}\boldsymbol{P}(t)$. From $\rho(0)=\rho(1)=0$ and the fact that $\boldsymbol{P}$ is a shortest path we conclude $I_\Delta(\boldsymbol{Q}_\rho) \geq I_\Delta(\boldsymbol{P}) = D_\Delta(\boldsymbol{P}(0)^{-1}\boldsymbol{P}(1))$.

Moreover, we find $\mathrm{D}L^{-1}_{\boldsymbol{Q}_\rho}\dot{\boldsymbol{Q}} = \mathrm{D}L^{-1}_{\boldsymbol{P}(t)}(\dot\rho(t)\boldsymbol{\xi}\boldsymbol{P}(t)+\dot{\boldsymbol{P}}(t)) = \dot\rho\boldsymbol{\kappa}(t)+\Xi(t)$ with $\boldsymbol{\kappa}(t) = \mathrm{Ad}^{-1}_{\boldsymbol{P}(t)}\boldsymbol{\xi}$. For fixed ρ consider the functional $K_\rho : \boldsymbol{g}\to[0,\infty)$ with

$$K_\rho(\boldsymbol{\xi}) := \textstyle\int_0^1 \Delta(\dot\rho(t)\mathrm{Ad}^{-1}_{\boldsymbol{P}(t)}\boldsymbol{\xi}+\Xi(t))\,\mathrm{d}t. \tag{5.11}$$

This functional is convex and by our assumption it attains the global minimum in $\boldsymbol{\xi}=0$. This implies $0\in\partial K_\rho(0)\subset\boldsymbol{g}^*$.

Let $f(t,\boldsymbol{\xi}) = \Delta(\Xi(t)+\dot\rho(t)\mathrm{Ad}^{-1}_{\boldsymbol{P}(t)}\boldsymbol{\xi})$ then $\partial f(t,\boldsymbol{\xi}) = \dot\rho(t)\mathrm{Ad}^{-*}_{\boldsymbol{P}(t)}[\partial\Delta(\Xi(t))]$ and Lemma A.1 below gives

$$\partial K_\rho(0) = \{\textstyle\int_0^1 \dot\rho(t)\boldsymbol{\eta}(t)\,\mathrm{d}t : \boldsymbol{\eta}(t)\in\mathrm{Ad}^*_{\boldsymbol{P}(t)^{-1}}[\partial\Delta(\Xi(t))] \text{ a.e.}\}.$$

From $0 \in \partial K_\rho(0)$ for all $\rho \in \mathrm{W}_0^{1,\infty}((0,1),\mathbb{R})$ and Lemma A.2 below we conclude that

$$E = \mathrm{ess} \bigcap_{t\in(0,1)} \mathrm{Ad}^{-*}_{\boldsymbol{P}(t)}[\partial\Delta(\Xi(t))] \subset \boldsymbol{g}^*$$

is nonempty. However, for any $\boldsymbol{\eta} \in E$ equation (5.10) is satisfied as $\Xi = \mathrm{D}L_{\boldsymbol{P}^{-1}}\dot{\boldsymbol{P}}$.

Clearly, $\boldsymbol{\eta} = 0$ would imply $\dot{\boldsymbol{P}}(t) = 0$ for a.e. $t \in [0,1]$. Hence, we conclude $\boldsymbol{\eta} \neq 0$. $\square$

This theorem provides us with a sufficient condition for the differentiability of shortest paths and hence of geodesic curves. If the boundary $\partial\mathbb{Q}$ of $\mathbb{Q}$ is of class C^1, then the normal cones $\mathrm{N}_{\boldsymbol{\eta}}\mathbb{Q}$ are one–dimensional rays $\{\lambda N(\boldsymbol{\eta}) : \lambda \geq 0\}$, where $N : \partial\mathbb{Q} \to \{\boldsymbol{\xi} \in \boldsymbol{g} : \|\boldsymbol{\xi}\| = 1\}$ is continuous. Thus, up to time reparametrization (5.10) is equivalent to $\dot{\boldsymbol{P}} = \mathrm{D}L_{\boldsymbol{P}} N(\mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta})$, whose solutions are continuously differentiable in time.

Corollary 5.5 *Let $\partial\mathbb{Q}$ be of class C^1 (or, equivalently, let $\{\, \boldsymbol{\xi} \in \boldsymbol{g} : \Delta(\boldsymbol{\xi}) \leq 1 \,\}$ be strictly convex), then all geodesic curves are C^1 functions in time.*

Example 5.6 To see the impact of this result we return to Example 5.1, where the abelian Lie group $G = \{\mathrm{diag}(\alpha,\beta) \in \mathrm{GL}(2) : \alpha,\beta > 0\}$ is considered with $\Delta(\boldsymbol{\xi}) = |\xi_1|+|\xi_2|$ where $\boldsymbol{\xi} = (\xi_1,\xi_2) \in \boldsymbol{g} = \mathbb{R}^2$.

We have $\partial\Delta(\boldsymbol{\xi}) = (\mathrm{Sign}(\xi_1), \mathrm{Sign}(\xi_2))$ where $\mathrm{Sign}(\xi_1) = +1$, $[-1,1]$, and -1 for $\xi_1 > 0$, $\xi_1 = 0$, and $\xi_1 < 0$, respectively. Using $Ad^*_{\boldsymbol{P}}(\boldsymbol{\eta}) = \boldsymbol{\eta}$ as G is abelian equation, (5.10) takes the form

$$\boldsymbol{\eta} = \binom{\eta_1}{\eta_2} \in \binom{\mathrm{Sign}(\dot{\alpha}(t)/\alpha(t))}{\mathrm{Sign}(\dot{\beta}(t)/\beta(t))} \text{ for fixed } \boldsymbol{\eta} \in \mathbb{R}^2.$$

This clearly holds if and only if $\dot{\alpha}$ and $\dot{\beta}$ do not change sign. Then only the cases $\boldsymbol{\eta} = (\pm 1, \pm 1)$ are relevant.

We might consider solutions of (5.10) as **generalized geodesic curves**. It is unclear whether they are real geodesic curves since the usual argument in Riemannian geometry involves strict convexity of $\widehat{D}(\boldsymbol{P}, \dot{\boldsymbol{P}})$ in $\dot{\boldsymbol{P}}$ which is certainly not the case here. The next example shows that we have to expect nonsmooth curves as soon as $\Delta : \boldsymbol{g}\backslash\{\mathbf{0}\} \to [0,\infty)$ is not differentiable.

Example 5.7 *Consider $G = \mathrm{GL}(d)$ and $\Delta(\boldsymbol{\xi}) = \|\boldsymbol{\xi}_{\mathrm{sym}}\|+\|\boldsymbol{\xi}_{\mathrm{anti}}\|$ such that $\mathbb{Q} = \{\boldsymbol{\eta}_{\mathrm{sym}}+\boldsymbol{\eta}_{\mathrm{anti}} : \|\boldsymbol{\eta}_{\mathrm{sym}}\| \leq 1, \|\boldsymbol{\eta}_{\mathrm{anti}}\| \leq 1\}$. Using the Sign function with $\mathrm{Sign}(\mathbf{0}) = \{\widehat{\boldsymbol{\xi}} : \|\widehat{\boldsymbol{\xi}}\| \leq 1\}$ and $\mathrm{Sign}(\boldsymbol{\xi}) = \{\frac{1}{\|\boldsymbol{\xi}\|}\boldsymbol{\xi}\}$ for $\boldsymbol{\xi} \neq \mathbf{0}$, we find the formula*

$$\partial\Delta(\boldsymbol{\xi}) = \mathrm{Sign}(\boldsymbol{\xi}_{\mathrm{sym}}) + \mathrm{Sign}(\boldsymbol{\xi}_{\mathrm{anti}}).$$

Now define the nonsmooth curve $\boldsymbol{P} : [-1,1] \to \mathrm{GL}(d)$ via

$$\boldsymbol{P}(t) = \begin{cases} \boldsymbol{Q}(t)\boldsymbol{\Omega}(t) = \mathrm{e}^{(\boldsymbol{\sigma}-\boldsymbol{\omega})t}\mathrm{e}^{\boldsymbol{\omega}t} & \textit{for} \quad t \leq 0, \\ \boldsymbol{\Omega}(t) = \mathrm{e}^{\boldsymbol{\omega}t} & \textit{for} \quad t \geq 0, \end{cases}$$

with $\boldsymbol{\omega} = -\boldsymbol{\omega}^\top, \boldsymbol{\sigma} = \boldsymbol{\sigma}^\top$ and $\|\boldsymbol{\omega}\| = \|\boldsymbol{\sigma}\| = 1$. For $t < 0$ we find

$$\partial\Delta(\mathrm{D}L^{-1}_{\boldsymbol{P}}\dot{\boldsymbol{P}}) = \partial\Delta(\mathrm{Ad}_{\boldsymbol{\Omega}(t)}\boldsymbol{\sigma}) = \mathrm{Ad}^*_{\boldsymbol{\Omega}(t)}\boldsymbol{\sigma}+(\mathrm{Sign}(\mathbf{0}))_{\mathrm{anti}},$$

and for $t>0$ we have

$$\partial\Delta(\mathrm{D}L_{\boldsymbol{P}}^{-1}\dot{\boldsymbol{P}}) = \partial\Delta(\boldsymbol{\omega}) = (\mathrm{Sign}(\boldsymbol{0}))_{\mathrm{sym}}+\boldsymbol{\omega}.$$

For the left–hand side in (5.10) we obtain, with $\boldsymbol{\eta}=\boldsymbol{\sigma}+\boldsymbol{\omega}$, for $t\leq 0$ the relation

$$\mathrm{Ad}^*_{\boldsymbol{P}(t)}\boldsymbol{\eta} = \mathrm{Ad}^*_{\boldsymbol{\Omega}(t)}[\mathrm{Ad}^*_{\boldsymbol{Q}(t)}(\boldsymbol{\sigma}+\boldsymbol{\omega})] = \mathrm{Ad}^*_{\boldsymbol{\Omega}(t)}[\boldsymbol{\sigma}+\boldsymbol{\omega}] = [\mathrm{Ad}^*_{\boldsymbol{\Omega}(t)}\boldsymbol{\sigma}]+\boldsymbol{\omega}$$

and for $t\geq 0$ we have $\mathrm{Ad}^*_{\boldsymbol{P}(t)}\boldsymbol{\eta} = \mathrm{Ad}^*_{\boldsymbol{\Omega}(t)}(\boldsymbol{\sigma}+\boldsymbol{\omega}) = \mathrm{Ad}^*_{\boldsymbol{\Omega}(t)}\boldsymbol{\sigma}+\boldsymbol{\omega}$. Hence, we see that (5.10) is satisfied for $t\in[-1,1]\backslash\{0\}$.

6 Geodesics for von Mises dissipation

In the case of von Mises plasticity one assumes isotropy of the dissipation functional on $G=\mathrm{SL}(d)$. Moreover, it is assumed to be generated by a scalar product, see (4.1).

$$\Delta_{\mathbb{B}}(\boldsymbol{\xi}) = (\langle \mathbb{B}\boldsymbol{\xi},\boldsymbol{\xi}\rangle)^{1/2}.$$

The geodesic curves can be obtained in two ways. On the one hand we can use the theory of Riemannian metrics where it is easy to show that the geodesic curves are obtained also by minimizing the energy $\int_0^1 \frac{1}{2}[\Delta(\mathrm{D}L^{-1}_{\boldsymbol{P}(t)}\dot{\boldsymbol{P}}(t))]^2\,\mathrm{d}t$. Then our smooth theory starting with (5.5) applies. For a more general Lie group approach we refer to [Mit00a].

On the other hand we can specialize the general nonsmooth theory to this situation. We have

$$\begin{aligned}\partial\Delta_{\mathbb{B}}(0) &= \{\boldsymbol{\eta}\in \mathrm{sl}(d) : \langle\boldsymbol{\eta},\mathbb{B}^{-1}\boldsymbol{\eta}\rangle\leq 1\}\\ \text{and } \partial\Delta_{\mathbb{B}}(\xi) &= \{\tfrac{1}{\langle\mathbb{B}\boldsymbol{\xi},\boldsymbol{\xi}\rangle^{1/2}}\mathbb{B}\boldsymbol{\xi}\} \text{ for } \boldsymbol{\xi}\neq 0.\end{aligned}$$

Thus, under the assumption $\dot{\boldsymbol{P}}\neq 0$ the half–integrated equation (5.8) takes the form

$$\dot{\boldsymbol{P}} = \lambda(t)\,\mathrm{D}L_{\boldsymbol{P}}\mathbb{B}^{-1}(\mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta}), \tag{6.1}$$

where $\boldsymbol{\eta}\in\boldsymbol{g}^*$ is fixed and $\lambda(t)$ is a positive scalar fixing the time rate. Without loss of generality we take $\lambda\equiv 1$ further on. In matrix notation for $G=\mathrm{GL}(d)$ or $G=\mathrm{SL}(d)$ we have $\dot{\boldsymbol{P}} = \boldsymbol{P}\mathbb{B}^{-1}(\boldsymbol{P}^\top\boldsymbol{\eta}\boldsymbol{P}^{-\top})$.

In order to solve (6.1) we need additional information. In von Mises plasticity we can use the isotropy which manifests in the material symmetry group $S=\mathrm{SO}(d)$ through $\Delta_{\mathbb{B}}(\mathrm{Ad}_{\boldsymbol{R}}\boldsymbol{\xi}) = \Delta_{\mathbb{B}}(\xi)$. For $\mathbb{B}$ this is equivalent to

$$\mathbb{B}^{-1}\mathrm{Ad}^*_{\boldsymbol{R}} = \mathrm{Ad}_{\boldsymbol{R}^{-1}}\mathbb{B}^{-1} \text{ for all } \boldsymbol{R}\in\mathrm{SO}(d). \tag{6.2}$$

In the case $G=\mathrm{SL}(d)$ this leads, according to (4.2), to

$$\mathbb{B}^{-1}\boldsymbol{\eta} = \tfrac{1}{\gamma_2}\boldsymbol{\eta}_{\mathrm{sym}} + \tfrac{1}{\gamma_3}\boldsymbol{\eta}_{\mathrm{anti}} = \tfrac{1}{2\gamma_2}(\boldsymbol{\eta}+\boldsymbol{\eta}^\top) + \tfrac{1}{2\gamma_3}(\boldsymbol{\eta}-\boldsymbol{\eta}^\top) \tag{6.3}$$

with $\gamma_2,\gamma_3>0$.

Theorem 6.1 *If $\Delta_{\mathbb{B}}$ is isotropic (see (6.2) or (6.3)) then the unique solution of (6.1) with $\lambda\equiv 1$ and $\boldsymbol{P}(0)=\boldsymbol{I}$ is given by*

$$\boldsymbol{P}(t) = \mathrm{e}^{t\boldsymbol{\alpha}}e^{t\boldsymbol{\omega}} \textit{ with } \boldsymbol{\alpha} = \tfrac{1}{\gamma_2}\boldsymbol{\eta}^\top \textit{ and } \boldsymbol{\omega} = (\tfrac{1}{\gamma_2}+\tfrac{1}{\gamma_3})\boldsymbol{\eta}_{\mathrm{anti}}. \tag{6.4}$$

PROOF: We make the ansatz $\boldsymbol{P}(t)=\mathrm{e}^{t\boldsymbol{\alpha}}\mathrm{e}^{t\boldsymbol{\omega}}$ with $\boldsymbol{\omega}=-\boldsymbol{\omega}^\top\in\mathrm{so}(d)$. With $\mathrm{Ad}^*_{\boldsymbol{P}_1\boldsymbol{P}_2}=\mathrm{Ad}^*_{\boldsymbol{P}_2}\mathrm{Ad}^*_{\boldsymbol{P}_1}$ we obtain $\dot{\boldsymbol{P}}=\mathrm{e}^{t\boldsymbol{\alpha}}(\boldsymbol{\alpha}+\boldsymbol{\omega})\mathrm{e}^{t\boldsymbol{\omega}}$ and

$$\boldsymbol{P}\mathbb{B}^{-1}[\mathrm{Ad}^*_{\boldsymbol{P}}\boldsymbol{\eta}]=\boldsymbol{P}\mathrm{Ad}_{\mathrm{e}^{-t\boldsymbol{\omega}}}(\mathbb{B}^{-1}[\mathrm{Ad}^*_{\mathrm{e}^{t\boldsymbol{\alpha}}}\boldsymbol{\eta}])=\mathrm{e}^{t\boldsymbol{\alpha}}(\mathbb{B}^{-1}[\mathrm{e}^{t\boldsymbol{\alpha}^\top}\boldsymbol{\eta}\mathrm{e}^{-t\boldsymbol{\alpha}^\top}])\mathrm{e}^{t\boldsymbol{\omega}}.$$

Clearly, we have found a solution if

$$\boldsymbol{\alpha}^\top\boldsymbol{\eta}=\boldsymbol{\eta}\boldsymbol{\alpha}^\top \text{ and } \boldsymbol{\alpha}+\boldsymbol{\omega}=\mathbb{B}^{-1}\boldsymbol{\eta}$$

holds. The second equation gives $\boldsymbol{\alpha}_{\mathrm{sym}}=\frac{1}{\gamma_2}\boldsymbol{\eta}_{\mathrm{sym}}$ and $\boldsymbol{\alpha}_{\mathrm{anti}}=\frac{1}{\gamma_3}\boldsymbol{\eta}_{\mathrm{anti}}-\boldsymbol{\omega}$. This reduces the first equation to $\boldsymbol{\eta}\boldsymbol{\omega}-\boldsymbol{\omega}\boldsymbol{\eta}=(\frac{1}{\gamma_2}+\frac{1}{\gamma_3})(\boldsymbol{\eta}_{\mathrm{sym}}\boldsymbol{\eta}_{\mathrm{anti}}-\boldsymbol{\eta}_{\mathrm{anti}}\boldsymbol{\eta}_{\mathrm{sym}})$ and we find the desired result. ☐

Corollary 6.2 *Under the above assumptions we have* $D_{\mathbb{B}}(\boldsymbol{R}^\top\boldsymbol{P}\boldsymbol{R})=D_{\mathbb{B}}(\boldsymbol{P})$ *and*

$$\begin{aligned}&D_{\mathbb{B}}(\boldsymbol{P})=\min\{\langle\mathbb{B}\boldsymbol{\xi},\boldsymbol{\xi}\rangle^{1/2}:\boldsymbol{P}=M_\delta(\boldsymbol{\xi})\}\\ &\text{with } \delta=\gamma_3/\gamma_2 \text{ and } M_\delta(\boldsymbol{\xi})=\left\{\begin{array}{ccc}\mathrm{sl}(d)&\to&\mathrm{SL}(d);\\ \boldsymbol{\xi}&\mapsto&\mathrm{e}^{\boldsymbol{\xi}_{\mathrm{sym}}-\delta\boldsymbol{\xi}_{\mathrm{anti}}}\mathrm{e}^{(1+\delta)\boldsymbol{\xi}_{\mathrm{anti}}}.\end{array}\right.\end{aligned}\tag{6.5}$$

Here M_δ is a surjective map for $\delta>0$. In particular, the limit $\gamma_3\searrow 0$ (plastic spin is dissipationless) leads to the formula

$$D_{\mathbb{B}_0}(\boldsymbol{P})=\gamma_2\|\log(\boldsymbol{P}\boldsymbol{P}^\top)^{1/2}\|_F.$$

PROOF: For $\boldsymbol{\xi}\in\mathrm{sl}(d)$ let $\boldsymbol{P}(t)=M_\delta(t\boldsymbol{\xi})$, then (6.4) holds with $\boldsymbol{\alpha}=\boldsymbol{\xi}_{\mathrm{sym}}-\delta\boldsymbol{\xi}_{\mathrm{anti}}$ and $\boldsymbol{\omega}=(1+\delta)\boldsymbol{\xi}_{\mathrm{anti}}$. Moreover, we have $\boldsymbol{P}^{-1}\dot{\boldsymbol{P}}=\mathrm{e}^{-t\boldsymbol{\omega}}(\boldsymbol{\alpha}+\boldsymbol{\omega})\mathrm{e}^{t\boldsymbol{\omega}}$. Hence $I_{\Delta_{\mathbb{B}}}(\boldsymbol{P}(\cdot))=\Delta_{\mathbb{B}}(\boldsymbol{\alpha}+\boldsymbol{\omega})=(\gamma_2\|\boldsymbol{\xi}_{\mathrm{sym}}\|_F^2+\gamma_3\|\boldsymbol{\xi}_{\mathrm{anti}}\|_F^2)^{1/2}$. As we know by Theorem 5.3 there exists a shortest path. By the above this path has to have the form $\boldsymbol{P}(t)=M_\delta(t\boldsymbol{\xi})$ and thus (6.5) follows.

The special case $\delta=0$ follows since $M_\delta(\boldsymbol{\xi})$ depends continuously on $\delta\in\mathbb{R}$ with $M_0(\boldsymbol{\xi})=\mathrm{e}^{\boldsymbol{\xi}_{\mathrm{sym}}}\mathrm{e}^{\boldsymbol{\xi}_{\mathrm{anti}}}$. Thus, the equation $\boldsymbol{P}=\mathrm{e}^{\boldsymbol{\xi}_{\mathrm{sym}}}\mathrm{e}^{\boldsymbol{\xi}_{\mathrm{anti}}}$ has a unique solution for $\boldsymbol{\xi}_{\mathrm{sym}}$ from $\boldsymbol{P}\boldsymbol{P}^\top=\mathrm{e}^{2\boldsymbol{\xi}_{\mathrm{sym}}}$. (We use the unique decomposition $\boldsymbol{P}=\boldsymbol{V}\boldsymbol{R}$ with $\boldsymbol{V}=\boldsymbol{V}^\top>0$ and $\boldsymbol{R}\in\mathrm{SO}(d)$, also called Cartan decomposition.) Here $\boldsymbol{\xi}_{\mathrm{anti}}$ is unique up to solutions of $\mathrm{e}^{\boldsymbol{\xi}}=\boldsymbol{I}$. ☐

In many crystal experiments the plastic spin appears to be very small. Hence, the case $\gamma_3\nearrow\infty$ should be studied since $\gamma_3\gg\gamma_2$ makes plastic spin less favorable than plastic stretching. The above distance metric can be rewritten as

$$D_{\gamma_2,\gamma_3}(\boldsymbol{P})=\min\{(\gamma_2\|\boldsymbol{\sigma}\|^2+\tfrac{\gamma_2^2}{\gamma_3}\|\boldsymbol{\omega}\|^2)^{1/2}:\boldsymbol{\sigma}=\boldsymbol{\sigma}^\top,\boldsymbol{\omega}=-\boldsymbol{\omega}^\top,\boldsymbol{P}=\mathrm{e}^{\boldsymbol{\sigma}-\boldsymbol{\omega}}\mathrm{e}^{(1+1/\delta)\boldsymbol{\omega}}\}$$

and we conjecture that the limit for $\gamma_3\to 0$ is given by

$$D_{\mathrm{no\,spin}}(\boldsymbol{P})=\inf\{\gamma_2^{1/2}\|\boldsymbol{\sigma}\|:\boldsymbol{\sigma}=\boldsymbol{\sigma}^\top,\boldsymbol{\omega}=-\boldsymbol{\omega}^\top,\boldsymbol{P}=\mathrm{e}^{\boldsymbol{\sigma}-\boldsymbol{\omega}}\mathrm{e}^{\boldsymbol{\omega}}\}.$$

In case the matrix $\boldsymbol{P}$ has no representation in the desired form the infimum is taken to be $+\infty$.

Finally we return to the Hamiltonian formulation for the geodesic curves. We recall the transformed Hamiltonian system (5.7) with $H(\boldsymbol{\eta}) = \frac{1}{2}\langle \mathbb{B}^{-1}\boldsymbol{\eta}, \boldsymbol{\eta}\rangle$

$$\begin{aligned} \dot{\boldsymbol{P}} &= \mathrm{D}L_{\boldsymbol{P}}\mathrm{D}H^0(\boldsymbol{\eta}) = \mathrm{D}L_{\boldsymbol{P}}(\mathbb{B}^{-1}\boldsymbol{\eta}). \\ \dot{\boldsymbol{\eta}} &= J_{\mathrm{LP}}(\boldsymbol{\eta})\,\mathrm{D}H(\boldsymbol{\eta}) = [\![\, (\mathbb{B}^{-1}\boldsymbol{\eta})^{\mathrm{T}}\boldsymbol{\eta} - \boldsymbol{\eta}(\mathbb{B}^{-1}\boldsymbol{\eta})^{\top} \,]\!] \,. \end{aligned} \tag{6.6}$$

In our case the Lie–Poisson structure $J_{\mathrm{LP}} : \boldsymbol{g}^* \to \mathrm{Skew}(\boldsymbol{g}, \boldsymbol{g}^*)$ in matrix notation reads $J_{\mathrm{LP}}(\boldsymbol{\eta})\boldsymbol{\xi} = \boldsymbol{\xi}^{\top}\boldsymbol{\eta} - \boldsymbol{\eta}\boldsymbol{\xi}^{\top}$. Because of the isotropy (6.3) the second equation in (6.6) decouples as follows:

$$\begin{aligned} \dot{\boldsymbol{\eta}}_{\mathrm{sym}} &= \gamma(\boldsymbol{\eta}_{\mathrm{sym}}\boldsymbol{\eta}_{\mathrm{anti}} - \boldsymbol{\eta}_{\mathrm{anti}}\boldsymbol{\eta}_{\mathrm{sym}}), \\ \dot{\boldsymbol{\eta}}_{\mathrm{anti}} &= 0, \quad \text{where } \gamma = \tfrac{1}{\gamma_2} + \tfrac{1}{\gamma_3}. \end{aligned} \tag{6.7}$$

This system can be solved explicitly by

$$\boldsymbol{\eta}(t) = \boldsymbol{\eta}_{\mathrm{sym}}(t) + \boldsymbol{\eta}_{\mathrm{anti}}(t) = \mathrm{e}^{-t\gamma\boldsymbol{\omega}}\boldsymbol{\sigma}e^{t\gamma\boldsymbol{\omega}} + \boldsymbol{\omega}$$

where $\boldsymbol{\omega} = -\boldsymbol{\omega}^{\top}$ and $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\top}$. From this and the first equation in (6.6) we find $\boldsymbol{P}(t)$ via the ansatz $\boldsymbol{P}(t) = \widehat{\boldsymbol{P}}(t)\mathrm{e}^{t\gamma\boldsymbol{\omega}}$ which leads to $\dot{\widehat{\boldsymbol{P}}} = \widehat{\boldsymbol{P}}\frac{1}{\gamma_2}(\boldsymbol{\sigma} - \boldsymbol{\omega})$. This justifies the ansatz in Theorem 6.1 and shows that there are no other possibilities.

A Set–valued calculus of variations

In this appendix we provide the two lemmas which were needed to prove Theorem 5.4. In the first lemma we give a formula for the subgradient of a function which is defined as an integral over convex functions.

Lemma A.1 *Let Y be a Banach space and Y^* its dual. Let $J(y) = \int_0^1 f(t,y)\,\mathrm{d}t$ where $f : [0,1] \times Y \to [0,\infty)$ is measurable and $f(t,\cdot)$ is convex with compact $\partial f(t,y)$ for each t. Then, $J : Y \to [0,\infty]$ is convex and*

$$\partial J(\eta) = \{\textstyle\int_0^1 w(t)\,\mathrm{d}t \in Y^* : w \in \mathrm{L}^1((0,1), Y^*),\ w(t) \in \partial f(t,\eta) \text{ for a.e. } t \in [0,1]\}. \tag{A.1}$$

Proof: Fix $y \in Y$ and denote the right–hand side in (A.1) by M. Take any w as defined in M. Then, $f(t, y{+}z) \geq f(t,y) + \langle w(t), z\rangle$ for all z and a.e. t. By integration we conclude $\int_0^1 w(t)\,\mathrm{d}t \in M$ and $M \subset \partial J(\eta)$ is proved.

We assume $w \in \partial J(y)\backslash M$ and derive a contradiction.

We have $J(y{+}z) \geq J(y) + \langle w, z\rangle$ for all $z \in Y$. Since all $\partial f(t,y)$ are compact convex sets the set M is compact and convex as well. By the separation theorem there exist $\varepsilon > 0$ and $\hat{z} \in Y$ with $\langle w, \hat{z}\rangle \geq \varepsilon + \langle m, \hat{z}\rangle$ for all $m \in M$. Now consider $j(s) = J(y{+}s\hat{z}) - J(y)$. From $w \in \partial J(y)$ we conclude $j(s) \geq \langle w, \hat{z}\rangle s \geq s(\varepsilon + \langle m, \hat{z}\rangle)$ for $s \geq 0$.

The convexity of $f(t,\cdot)$ implies

$$f(t, y{+}s\hat{z}) = f(t,y) + s\beta(t) + o(s) \text{ with } \beta(t) = \langle \hat{w}(t), \hat{z}\rangle$$

for some $\hat{w}(t) \in \partial f(t,y)$. Hence

$$\begin{aligned} j(s) &= \textstyle\int_0^1 [f(t, y{+}s\hat{z}) - f(t,y)]\,\mathrm{d}s = s\int_0^1 \langle \hat{w}(t), \hat{z}\rangle\,\mathrm{d}s + o(s) \\ &= s\langle \widehat{m}, \hat{z}\rangle + o(s) \text{ where } \widehat{m} = \textstyle\int_0^1 \hat{w}(t)\,\mathrm{d}s \in M. \end{aligned}$$

This is the desired contradiction and $\partial J(y) \subset M$ is proved. $\square$

The second lemma generalizes the fundamental lemma in the calculus of variations which states that a function $a : (0,1) \to \mathbb{R}$ must be constant if it satisfies $\int_0^1 \dot\rho(t)a(t)\,\mathrm{d}t = 0$ for all $\rho \in \mathrm{W}_0^{1,\infty}((0,1))$.

Lemma A.2 *For each $t \in [0,1]$ let $C(t) \subset \mathbb{R}^n$ be a compact set such that the mapping $(t,a) \mapsto \mathcal{X}_{C(t)}(a)$ from $[0,1] \times \mathbb{R}^n$ into $[0,\infty]$ is measurable. Moreover assume that for each $\rho \in \mathrm{W}_0^{1,\infty}((0,1))$ we have*

$$0 \in K_\rho := \{\textstyle\int_0^1 \dot\rho(t)a(t)\,\mathrm{d}t \mid a : [0,1] \to \mathbb{R}^n \text{ measurable, } a(t) \in C(t) \text{ a.e.}\} \subset \mathbb{R}^n.$$

Then there exists a vector $b^* \in \mathbb{R}^n$ such that $b^* \in C(t)$ for a.e. $t \in [0,1]$.

Proof: We first prove the result for piecewise constant maps $t \mapsto C(t)$. Assume $[0,1] = \cup_{i=1}^n A_i$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ and A_i measurable and $C(t) = C_i$ for $t \in A_i$.

Then we have to show that $\cap_{i=1}^n C_i \neq \emptyset$ under the assumption $0 \in \sum_{i=1}^n \lambda_i C_i = \{\sum_{i=1}^n \lambda_i a_i : a_i \in C_i\}$ for all $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ with $\sum_{i=1}^n \lambda_i = 0$. We prove this by induction on n. For $n = 2$ we have $0 \in \lambda_1 C_1 + \lambda_2 C_2$ with $\lambda_1 + \lambda_2 = 0$ if and only if $0 \in C_1 - C_2 = \{a_1 - a_2 : a_j \in C_j\}$. But this is the same as $C_1 \cap C_2 \neq \emptyset$.

For the induction from n to $n+1$ let $B_n = \cap_{i=1}^n C_i$ which is nonempty by the induction hypothesis. We need to show $B_n \cap C_{n+1} \neq \emptyset$. First note the simple identity $B_n = \bigcap\{\sum_1^n \mu_i C_i : \mu_1 + \ldots + \mu_n = 1\}$. From $0 \in \sum_{i=1}^{n+1} \lambda_i C_i$ for all $\lambda_1, \ldots, \lambda_{n+1}$ with $\sum_{i=1}^{n+1} \lambda_i = 0$ we conclude

$$\begin{aligned} 0 \;\in\; & \bigcap\{(\textstyle\sum_{i=1}^n \mu_i C_i) - C_{n+1} : \sum_1^n \mu_i = 1\} \\ =\; & \{\textstyle\sum_{i=1}^n \mu_i C_i : \sum_1^n \mu_i = 1\} - C_{n+1} = B_n - C_{n+1}. \end{aligned}$$

But this exactly means $B_n \cap C_{n+1} = \bigcap_{i=1}^{n+1} C_i \neq \emptyset$. This proves the assertion for piecewise constant functions $(t \mapsto C(t))$.

The general case is obtained by discretization. For $m \in \mathbb{N}$ we define the subintervals $I_k^m = ((k-1)2^{-m}, k2^{-m})$ for $k = 1, \ldots, 2^m$ and the sets

$$C_k^m = \{\textstyle\int_0^1 a(t)\,\mathrm{d}t : a(t) \in C((k-1+t)2^{-m}) \text{ for a.e. } t \in [0,1]\}.$$

By taking all $\rho \in W_0^{1,\infty}((0,1))$ such that $\dot\rho$ is constant on I_k^m we have $0 \in \sum_{k=1}^{2^m} \lambda_k C_k^m$ for $\sum_k \lambda_k = 0$. From the above we conclude

$$B^m = \textstyle\bigcup_{k=1}^{2^m} C_k^m \neq \emptyset.$$

By construction we have $C_k^m = \frac{1}{2}C_{2k-1}^{m+1} + \frac{1}{2}C_{2k}^{m+1}$. Because of $B^{m+1} \subset C_i^{m+1}$ for all i we find $B^{m+1} \subset C_j^m$ for all j and finally $B^{m+1} \subset B^m$. Since all B^m are compact we know that $B^* = \cap_{m=1}^\infty B^m$ is nonempty and satisfies $B^* \subset B^m$ for all m. This implies $B^* \subset C_k^m$ for all $m \in \mathbb{N}$ and $k = 1, \ldots 2^m$. By the measurability of the family $(t \mapsto C(t))$ we conclude that for almost all t the sets $C_{[t2^m]}^m$ converge for $m \to \infty$ to $C(t)$. Hence $\emptyset \neq B^* \subset C(t)$ for a.e. $t \in [0,1]$. $\square$

References

[AbM78] R. Abraham, J.E. Marsden. *Foundations of Mechanics.* Addison–Wesley, 2nd ed., 1978.

[Bal77] J.M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Rat. Mech. Analysis* **63** (1977) 337–403.

[CHM00] C. Carstensen, K. Hackl, A. Mielke. Nonconvex potentials and microstructures in finite–strain plasticity. In preparation, June 2000.

[Dac89] B. Dacorogna. *Direct Methods in the Calculus of Variations.* Springer–Verlag 1989.

[GoM00] S. Govindjee, A. Mielke: The mixture function and flow rules for phase transformation problems. Universität Stuttgart, in preparation 2000.

[Gur00] M.E. Gurtin. On the plasticity of single crystals: free energy, microforces, plastic–strain gradients. *J. Mech. Physics Solids* **48** (2000) 989–1036.

[Hac97] K. Hackl: Generalized Standard Media and Variational Principles in Classical and Finite Strain Elastoplasticity. *J. Mech. Phys. Solids* **45** (1997) 667–688.

[HaR95] W. Han, B.D. Reddy. Computational plasticity: the variational basis and numerical analysis. *Comp. Mech. Advances* **2** (1995) 283–400.

[Lee69] E. H. Lee: Elastic-plastic deformation at finite strains *J. Appl. Mech.* **36** (1969) 1–6.

[MaH83] J.E. Marsden, T.J.R. Hughes. *Mathematical Foundations of Elasticity.* Prentice Hall 1983.

[Mie91] A. Mielke. *Hamiltonian and Lagrangian Flows on Center Manifolds*, Lecture Notes in Mathematics Vol. **1489**, Springer–Verlag, Berlin, 1991.

[MiS92] C. Miehe, E. Stein. A Canonical Model of Multiplicative Elasto-Plasticity. Formulation and Aspects of the Numerical Implementation *European J. Mechanics, A/Solids* **11** (1992) 25–43.

[MiT99] A. Mielke, F. Theil. A mathematical model for rate-independent phase transformations with hysteresis. In "*Models of Continuum Mechanics in Analysis and Engineering*, H.–D. Alber, R. Farwig (eds), Shaker Verlag, 1999", 117–129.

[Mit00a] D. Mittenhuber. Pseudo–Riemannian metrics on Lie groups. Preprint TU Darmstadt, Febr. 2000.

[Mit00b] D. Mittenhuber. Personal Communication by e–mail, August 2000.

[Mon93] M.D.P. Monteiro Marques. Differential Inclusions in Nonsmooth Mechanics (Shocks and Dry Friction). Birkhäuser, 1993.

[MTL00] A. Mielke, F. Theil, V.I. Levitas. A variational formulation of rate–independent phase transformations using an extremum principle. Submitted to *Archive Rat. Mech. Analysis* June 2000.

[OrR99] M. Ortiz, E.A. Repetto: Nonconvex energy minimisation and dislocation in ductile single crystals. *Journal of the Mechanics and Physics of Solids.* **47** (1999) 397–462.

[Sim88] J.C. Simo. A framework for finite strain elastoplasticity based on maximum plastic dissipation and the multiplicative decomposition. Part I: Continuum formulation. *Comp. Meth. Appl. Mech. Engrg.* **66** (1988) 199–219.

[SiO85] J.C. Simo, M. Ortiz: A Unified Approach to Finite Deformation Elastoplastic Analysis Based on the Use of Hyperelastic Constitutive Equations, *Comput. Meths. Appl. Mech. Engrg.* **49** (1985) 221–245.

[ZiW87] H. Ziegler, C. Wehrli : The Derivation of Constitutive Relations from the Free Energy and the Dissipation Function. In *Advances in Applied Mechanics 25*, Eds. T.Y. Wu & J.W. Hutchinson, Academic Press, New York 1987.