Journal of Convex Analysis
Volume 28 (2021), No. 1, 1–10

A Majorization Gradient Inequality for Symmetric Convex Functions

Marek Niezgoda
Department of Applied Mathematics and Computer Science
University of Life Sciences, 20-950 Lublin, Poland
marek.niezgoda@up.lublin.pl

Received: December 5, 2019
Accepted: December 15, 2019

We extend the standard two-points gradient inequality for a differentiable convex function to a four-points version satisfying Sherman's majorization condition.

Keywords: Convex function, gradient inequality, majorization, doubly stochastic matrix, Sherman's inequality.

2010 Mathematics Subject Classification: 26B25, 26D10, 52A41.

1. Preliminaries

In this expository section we collect some basic notation, definitions and facts.

As usual, $\mathbb{R}^n$ denotes the n-dimensional Euclidean space with inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i \quad \text{for } \mathbf{x} = (x_1, \dots, x_n)^T \text{ and } \mathbf{y} = (y_1, \dots, y_n)^T,$$

and *norm*
$$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \quad \text{for } \mathbf{x} = (x_1, \dots, x_n)^T.$$

Hereafter the symbol $(\cdot)^T$ stands for the *transpose* of a matrix.

The space $\mathbb{R}^n$ is equipped with the entrywise order $\leq$ given by

$$\mathbf{y} \leq \mathbf{x} \quad \text{iff} \quad y_i \leq x_i \quad \text{for each } i = 1, \dots, n,$$

where $\mathbf{x} = (x_1, \dots, x_n)^T$ and $\mathbf{y} = (y_1, \dots, y_n)^T$.

A function $\Phi \colon C \to \mathbb{R}$ defined on a convex set $C \subset \mathbb{R}^n$ is said to be *convex* on C, if for any points $\mathbf{x}_i \in C$ and scalars $t_i \geq 0$, $i = 1, \dots, k$, with $\sum_{i=1}^{k} t_i = 1$, the following *Jensen's inequality* holds:

$$\Phi \left(\sum_{i=1}^{k} t_i \mathbf{x}_i \right) \leq \sum_{i=1}^{k} t_i \Phi(\mathbf{x}_i). \tag{1}$$

It is well known that if $\Phi \colon C \to \mathbb{R}$ is a differentiable convex function then the *gradient inequality* holds, as follows:

$$\langle \nabla \Phi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq \Phi(\mathbf{x}) - \Phi(\mathbf{y}) \quad \text{for } \mathbf{x}, \mathbf{y} \in C, \tag{2}$$

where the symbol ∇ stands for the gradient, and $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{R}^n$.

ISSN 0944-6532 / $ 2.50

Given two n-tuples $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$, we say that $\mathbf{y}$ is *majorized* by $\mathbf{x}$, and write $\mathbf{y} \prec \mathbf{x}$, if

$$\sum_{i=1}^{l} y_{[i]} \leq \sum_{i=1}^{l} x_{[i]} \quad \text{for } l = 1, \ldots, n, \quad \text{and} \quad \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i,$$

where $x_{[1]} \geq \cdots \geq x_{[n]}$ and $y_{[1]} \geq \cdots \geq y_{[n]}$ are the entries of $\mathbf{x}$ and $\mathbf{y}$, respectively, stated in decreasing order [9, p. 8].

From now on the symbol conv means "the convex hull of", and $\mathbb{P}_n$ denotes the group of $n \times n$ permutation matrices.

It is known that
$$\mathbf{y} \prec \mathbf{x} \quad \text{if and only if} \quad \mathbf{y} \in \operatorname{conv} \mathbb{P}_n \mathbf{x} \tag{3}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (see [13], [9, p. 10]).

An $n \times m$ real matrix $\mathbf{S} = (s_{ij})$ is said to be *column stochastic* (resp. *row stochastic*) provided that $s_{ij} \geq 0$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$, and all column sums (resp. row sums) of $\mathbf{S}$ are equal to 1, i.e., $\sum_{i=1}^{n} s_{ij} = 1$ for $j = 1, \ldots, m$ (resp. $\sum_{j=1}^{m} s_{ij} = 1$ for $i = 1, \ldots, n$).

Observe that $\mathbf{S}$ is a column stochastic matrix iff $\mathbf{S}^T$ is a row stochastic matrix.

If an $n \times n$ real matrix $\mathbf{S}$ is both column stochastic and row stochastic then $\mathbf{S}$ is called *doubly stochastic* [9, pp. 29–30].

The set of all $n \times n$ doubly stochastic matrices is denoted by Ω_n.

By Birkhoff's Theorem,
$$\Omega_n = \operatorname{conv} \mathbb{P}_n \tag{4}$$

(see [9, Theorem A.2.]). In consequence, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, (3) can be rewritten as

$$\mathbf{y} \prec \mathbf{x} \ \text{ if and only if } \ \mathbf{y} = \mathbf{S}\mathbf{x} \tag{5}$$

for some doubly stochastic $n \times n$ matrix $\mathbf{S}$ depending on $\mathbf{x}$ and $\mathbf{y}$ [9, p. 33].

Following [9, pp. 79–154] we say that a function $\Phi \colon J^n \to \mathbb{R}$ with an interval $J \subset \mathbb{R}$ is *Schur-convex* (resp. *Schur-concave*) on J^n if for $\mathbf{x}, \mathbf{y} \in J^n$,

$$\mathbf{y} \prec \mathbf{x} \ \text{ implies } \ \Phi(\mathbf{y}) \leq (\text{resp.} \geq)\ \Phi(\mathbf{x}).$$

The following celebrated Majorization Theorem provides some important examples of Schur-convex functions.

Theorem A. (Hardy-Littlewood-Pólya [5] and Karamata [8])
Let $f \colon J \to \mathbb{R}$ be a real convex function defined on an interval $J \subset \mathbb{R}$.
Then, for $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in J^n$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in J^n$,

$$\mathbf{y} \prec \mathbf{x} \quad \textit{implies} \quad \sum_{i=1}^{n} f(y_i) \leq \sum_{i=1}^{n} f(x_i). \tag{6}$$

An essential extension of Theorem **A** is the following result.

Theorem B. (Sherman [14].) *Let f be a real convex function defined on an interval $J \subset \mathbb{R}$. Assume $\mathbf{a} = (a_1, \dots, a_m)^T \in \mathbb{R}_+^m$, $\mathbf{b} = (b_1, \dots, b_n)^T \in \mathbb{R}_+^n$, $\mathbf{x} = (x_1, \dots, x_m)^T \in J^m$ and $\mathbf{y} = (y_1, \dots, y_n)^T \in J^n$.*

If
$$\mathbf{y} = \mathbf{S}\mathbf{x} \quad \text{and} \quad \mathbf{a} = \mathbf{S}^T\mathbf{b} \tag{7}$$
for some $n \times m$ row stochastic matrix $\mathbf{S} = (s_{ij})$, then
$$\sum_{i=1}^{n} b_i f(y_i) \le \sum_{j=1}^{m} a_j f(x_j). \tag{8}$$

The interested reader is referred to [1, 2, 3, 4, 6, 7, 10, 11, 12] for interpretations, generalizations and applications of *Sherman's inequality* (8). Statement (7) is called *Sherman's condition.*

Note that the Sherman inequality (8) says that
$$\Phi_1(f(\mathbf{y}), \mathbf{b}) \le \Phi_2(f(\mathbf{x}), \mathbf{a}), \tag{9}$$
where $\Phi_1(\cdot,\cdot) = \langle \cdot,\cdot \rangle_1$ is the standard inner product on $\mathbb{R}^n$ and $\Phi_2(\cdot,\cdot) = \langle \cdot,\cdot \rangle_2$ is the standard inner product on $\mathbb{R}^m$, and $f(\mathbf{z}) = (f(z_1), \dots, f(z_k))^T$ for a vector $\mathbf{z} = (z_1, \dots, z_k)^T \in J^k$ and any $k \in \mathbb{N}$.

In what follows we develop this idea for another pair of (nonlinear) maps Φ_1 and Φ_2 induced by both the sides of the gradient inequality (2) for a differentiable convex function Φ.

See [12] for a general discussion of nonlinear Sherman type inequalities in the case $\Phi_1 = \Phi_2$.

2. An extension of the gradient inequality

Throughout the symbols J and J_0 represent any intervals in $\mathbb{R}$.

A function F defined on J^n is said to be $\mathbb{P}_n$-*invariant* (also called, *symmetric*), if
$$F(\mathbf{p}\mathbf{x}) = F(\mathbf{x}) \quad \text{for } \mathbf{p} \in \mathbb{P}_n \text{ and } \mathbf{x} \in J^n. \tag{10}$$

A function $F \colon J^n \to \mathbb{R}^n$ is said to be $\mathbb{P}_n$-*equivariant* if
$$F(\mathbf{p}\mathbf{x}) = \mathbf{p}F(\mathbf{x}) \quad \text{for } \mathbf{p} \in \mathbb{P}_n \text{ and } \mathbf{x} \in J^n. \tag{11}$$

Example 2.1. If $f \colon J \to \mathbb{R}$ is a one-variable real function, then its action can expand from J to J^n by
$$F\left((x_1, \dots, x_n)^T\right) = (f(x_1), \dots, f(x_n))^T \quad \text{for } x_1, \dots, x_n \in J. \tag{12}$$

It is not hard to check that F is $\mathbb{P}_n$-equivariant.

Example 2.2. Let $F\colon J^n \to \mathbb{R}$ be a differentiable Schur-convex function. Then F is necessary $\mathbb{P}_n$-invariant. In consequence, the gradient function $\nabla F(\cdot)\colon J^n \to \mathbb{R}^n$ is $\mathbb{P}_n$-equivariant.

Hereafter $\leq$ stands for the componentwise order on $\mathbb{R}^l$ with any $l \in \mathbb{N}$.
We say that a function $F\colon J^n \to \mathbb{R}^l$ is *convex* on J^n if

$$F\left(\sum_{i=1}^{k} t_i \mathbf{x}_i\right) \leq \sum_{i=1}^{k} t_i F\left(\mathbf{x}_i\right) \tag{13}$$

for any $k \in \mathbb{N}$, $\mathbf{x}_i \in J^n$, $t_i \geq 0$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} t_i = 1$.

In the forthcoming theorem we present a majorization generalization of the gradient inequality (2) for a symmetric convex function Φ. Thus we obtain an inequality of the form (9) with the nonlinear maps Φ_1 and Φ_2 defined by (17)–(18).

Theorem 2.3. *Let $\Phi\colon J^n \to \mathbb{R}$ be a differentiable $\mathbb{P}_n$-invariant convex function, and $F\colon J_0^n \to J^n$ be a $\mathbb{P}_n$-equivariant convex function. Let $\mathbf{x}, \mathbf{y} \in J_0^n$ and $\mathbf{a}, \mathbf{b} \in J^n$ with $\nabla\Phi(\mathbf{b}) \geq 0$.*
If

$$\mathbf{y} = \mathbf{S}\mathbf{x} \quad \textit{and} \quad \mathbf{a} = \mathbf{S}^T\mathbf{b} \tag{14}$$

for some $n \times n$ doubly stochastic matrix $\mathbf{S}$, then

$$\langle \nabla\Phi(\mathbf{b}), F(\mathbf{y}) - \mathbf{b}\rangle \leq \Phi(F(\mathbf{x})) - \Phi(\mathbf{a}). \tag{15}$$

Remark 2.4. Inequality (2) gives the standard inequality:

$$\langle \nabla\Phi(\mathbf{b}), F(\mathbf{y}) - \mathbf{b}\rangle \leq \Phi(F(\mathbf{y})) - \Phi(\mathbf{b}). \tag{16}$$

Inequality (15) has a potential to be a refinement of (16) in some situations. Namely, in Theorem 2.3, if in addition the composite function $\Phi \circ F$ is Schur-concave and $\mathbf{b} = \mathbf{e} = (1, \ldots, 1)^T \in \mathbb{R}^n$, then (14) implies $\mathbf{y} \prec \mathbf{x}$ and $\mathbf{e} = \mathbf{S}^T\mathbf{e} = \mathbf{S}^T\mathbf{b} = \mathbf{a}$, so $\Phi(F(\mathbf{x})) \leq \Phi(F(\mathbf{y}))$ and $\Phi(\mathbf{b}) = \Phi(\mathbf{e}) = \Phi(\mathbf{a})$.

Proof of Theorem 2.3. We define

$$\Phi_1(\mathbf{v}, \mathbf{w}) = \langle \nabla\Phi(\mathbf{w}), \mathbf{v} - \mathbf{w}\rangle \quad \text{for } \mathbf{v}, \mathbf{w} \in J^n, \tag{17}$$

$$\Phi_2(\mathbf{v}, \mathbf{w}) = \Phi(\mathbf{v}) - \Phi(\mathbf{w}) \quad \text{for } \mathbf{v}, \mathbf{w} \in J^n. \tag{18}$$

(i) The maps Φ_1 and Φ_2 possess the following property

$$\Phi_1(\mathbf{p}\mathbf{v}, \mathbf{w}) \leq \Phi_2(\mathbf{v}, \mathbf{p}^T\mathbf{w}) \quad \text{for } \mathbf{v}, \mathbf{w} \in J^n \text{ and } \mathbf{p} \in \mathbb{P}_n. \tag{19}$$

To see this, we use the $\mathbb{P}_n$-invariance of Φ, and we get $\nabla\Phi(\mathbf{p}^T\mathbf{w}) = \mathbf{p}^T\nabla\Phi(\mathbf{w})$ in consequence. Therefore we obtain

$$\Phi_1(\mathbf{p}\mathbf{v}, \mathbf{w}) = \langle \nabla\Phi(\mathbf{w}), \mathbf{p}\mathbf{v} - \mathbf{w}\rangle = \langle \nabla\Phi(\mathbf{w}), \mathbf{p}(\mathbf{v} - \mathbf{p}^{-1}\mathbf{w})\rangle = \langle \mathbf{p}^T\nabla\Phi(\mathbf{w}), \mathbf{v} - \mathbf{p}^T\mathbf{w}\rangle$$
$$= \langle \nabla\Phi(\mathbf{p}^T\mathbf{w}), \mathbf{v} - \mathbf{p}^T\mathbf{w}\rangle \leq \Phi(\mathbf{v}) - \Phi(\mathbf{p}^T\mathbf{w}) = \Phi_2(\mathbf{v}, \mathbf{p}^T\mathbf{w}).$$

The last inequality is satisfied by (2). Thus (19) holds, as required.

(ii) For each $\mathbf{w} \in J^n$ the one-variable map $\Phi_1(\cdot, \mathbf{w})$ is convex on J^n, that is, for $\mathbf{v}_1, \ldots, \mathbf{v}_k \in J^n$, $t_1, \ldots, t_k \geq 0$, $\sum_{i=1}^k t_i = 1$,

$$\Phi_1\left(\sum_{i=1}^k t_i \mathbf{v}_i, \mathbf{w}\right) \leq \sum_{i=1}^k t_i \Phi_1(\mathbf{v}_i, \mathbf{w}). \tag{20}$$

In fact, we have

$$\begin{aligned}\Phi_1\left(\sum_{i=1}^k t_i \mathbf{v}_i, \mathbf{w}\right) &= \left\langle \nabla\Phi\left(\mathbf{w}\right), \sum_{i=1}^k t_i \mathbf{v}_i - \mathbf{w}\right\rangle = \left\langle \nabla\Phi\left(\mathbf{w}\right), \sum_{i=1}^k t_i\left(\mathbf{v}_i - \mathbf{w}\right)\right\rangle \\ &= \sum_{i=1}^k t_i \langle \nabla\Phi\left(\mathbf{w}\right), \mathbf{v}_i - \mathbf{w}\rangle = \sum_{i=1}^k t_i \Phi_1(\mathbf{v}_i, \mathbf{w}).\end{aligned}$$

This implies (20), as claimed.

(iii) For each $\mathbf{v} \in J^n$ the one-variable map $\Phi_2(\mathbf{v}, \cdot)$ is concave on J^n, that is, for $\mathbf{w}_1, \ldots, \mathbf{w}_k \in J^n$, $t_1, \ldots, t_k \geq 0$, $\sum_{i=1}^k t_i = 1$,

$$\sum_{i=1}^k t_i \Phi_2(\mathbf{v}, \mathbf{w}_i) \leq \Phi_2\left(\mathbf{v}, \sum_{i=1}^k t_i \mathbf{w}_i\right). \tag{21}$$

Indeed, since the function Φ is convex, we obtain

$$\begin{aligned}\sum_{i=1}^k t_i \Phi_2(\mathbf{v}, \mathbf{w}_i) &= \sum_{i=1}^k t_i\left(\Phi(\mathbf{v}) - \Phi(\mathbf{w}_i)\right) = \Phi(\mathbf{v}) - \sum_{i=1}^k t_i \Phi\left(\mathbf{w}_i\right) \\ &\leq \Phi(\mathbf{v}) - \Phi\left(\sum_{i=1}^k t_i \mathbf{w}_i\right) = \Phi_2\left(\mathbf{v}, \sum_{i=1}^k t_i \mathbf{w}_i\right).\end{aligned}$$

Therefore (21) holds. Furthermore, by Birkhoff's Theorem (see (4)), we have $\mathbf{S} = \sum_{i=1}^k t_i \mathbf{p}_i$ for some $\mathbf{p}_i \in \mathbb{P}_n$ and $t_i \geq 0$ with $\sum_{i=1}^k t_i = 1$.

By combining (20), (19), (21), with any $\mathbf{v} \in J^n$ and $\mathbf{v}_i = \mathbf{p}_i \mathbf{v}$ for $i = 1, \ldots, k$, we find that

$$\begin{aligned}\Phi_1\left(\sum_{i=1}^k t_i \mathbf{p}_i \mathbf{v}, \mathbf{b}\right) &\leq \sum_{i=1}^k t_i \Phi_1(\mathbf{p}_i \mathbf{v}, \mathbf{b}) \leq \sum_{i=1}^k t_i \Phi_2(\mathbf{v}, \mathbf{p}_i{}^T \mathbf{b}) \\ &\leq \Phi_2\left(\mathbf{v}, \sum_{i=1}^k t_i \mathbf{p}_i{}^T \mathbf{b}\right) = \Phi_2\left(\mathbf{v}, \mathbf{S}^T \mathbf{b}\right) = \Phi_2\left(\mathbf{v}, \mathbf{a}\right).\end{aligned} \tag{22}$$

The last equality is valid by (14). By putting $\mathbf{v} = F(\mathbf{x})$ into (22), we get

$$\Phi_1\left(\sum_{i=1}^k t_i \mathbf{p}_i F(\mathbf{x}), \mathbf{b}\right) \leq \Phi_2\left(F(\mathbf{x}), \mathbf{a}\right). \tag{23}$$

Because of the convexity (w.r.t. $\leq$) and $\mathbb{P}_n$-equivariance of F on J_0^n, we can write

$$F\left(\sum_{i=1}^{k} t_i \mathbf{p}_i \mathbf{x}\right) \leq \sum_{i=1}^{k} t_i F(\mathbf{p}_i \mathbf{x}) = \sum_{i=1}^{k} t_i \mathbf{p}_i F(\mathbf{x}). \tag{24}$$

Therefore,
$$F\left(\sum_{i=1}^{k} t_i \mathbf{p}_i \mathbf{x}\right) - \mathbf{b} \leq \sum_{i=1}^{k} t_i \mathbf{p}_i F(\mathbf{x}) - \mathbf{b}, \tag{25}$$

and further, as $\nabla\Phi(\mathbf{b}) \geq 0$, we have

$$\langle \nabla\Phi(\mathbf{b}), F\left(\sum_{i=1}^{k} t_i \mathbf{p}_i \mathbf{x}\right) - \mathbf{b}\rangle \leq \langle \nabla\Phi(\mathbf{b}), \sum_{i=1}^{k} t_i \mathbf{p}_i F(\mathbf{x}) - \mathbf{b}\rangle. \tag{26}$$

That is,
$$\Phi_1\left(F\left(\sum_{i=1}^{k} t_i \mathbf{p}_i \mathbf{x}\right), \mathbf{b}\right) \leq \Phi_1\left(\sum_{i=1}^{k} t_i \mathbf{p}_i F(\mathbf{x}), \mathbf{b}\right). \tag{27}$$

In light of (14), (27) and (23) we obtain

$$\begin{aligned} \Phi_1\left(F\left(\mathbf{y}\right), \mathbf{b}\right) &= \Phi_1\left(F(\mathbf{S}\mathbf{x}), \mathbf{b}\right) = \Phi_1\left(F\left(\sum_{i=1}^{k} t_i \mathbf{p}_i \mathbf{x}\right), \mathbf{b}\right) \\ &\leq \Phi_1\left(\sum_{i=1}^{k} t_i \mathbf{p}_i F(\mathbf{x}), \mathbf{b}\right) \leq \Phi_2\left(F(\mathbf{x}), \mathbf{a}\right). \end{aligned} \tag{28}$$

Because $\Phi_1\left(F\left(\mathbf{y}\right), \mathbf{b}\right) = \langle \nabla\Phi(\mathbf{b}), F(\mathbf{y}) - \mathbf{b}\rangle$ and $\Phi_2(F(\mathbf{x}), \mathbf{a}) = \Phi(F(\mathbf{x})) - \Phi(\mathbf{a})$ (see (17)–(18)), the statement (28) leads to

$$\langle \nabla\Phi(\mathbf{b}), F(\mathbf{y}) - \mathbf{b}\rangle \leq \Phi(F(\mathbf{x})) - \Phi(\mathbf{a}),$$

completing the proof of (15). □

Remark 2.5. In Theorem 2.3, if $F = \mathrm{id}$ (the identity map) with $J_0 = J$, then the assumption $\nabla\Phi(\mathbf{b}) \geq 0$ can be deleted, because in this case the needed inequalities (26) and (27) become equalities and hold trivially.

Remark 2.6. In Theorem 2.3, if $\mathbf{S} = I_n$(the $n \times n$ identity matrix) then $\mathbf{a} = \mathbf{b}$ and $\mathbf{y} = \mathbf{x}$ (see (14)). In consequence, if in addition $F = \mathrm{id}$ then inequality (15) reduces to the usual gradient inequality (2) for a differentiable convex function Φ:

$$\langle \nabla\Phi(\mathbf{a}), \mathbf{x} - \mathbf{a}\rangle \leq \Phi(\mathbf{x}) - \Phi(\mathbf{a}). \tag{29}$$

Therefore inequality (15) is an extension of the standard inequality (29).

Corollary 2.7. *Let $F\colon J^n \to \mathbb{R}$ be a Schur-convex function. If the function $\nabla F(\cdot)\colon J_0^n \to J^n$ is convex, then, under the remaining assumptions of Theorem 2.3 on Φ, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{a}$, $\mathbf{b}$, the following inequality holds true:*

$$\langle \nabla\Phi(\mathbf{b}), \nabla F(\mathbf{y}) - \mathbf{b}\rangle \leq \Phi(\nabla F(\mathbf{x})) - \Phi(\mathbf{a}). \tag{30}$$

Proof. Apply Theorem 2.3 and Example 2.2. □

We use the notation $\mathbf{e} = (1, \ldots, 1)^T \in \mathbb{R}^n$.

Corollary 2.8. *Let $\Phi\colon J^n \to \mathbb{R}$ be a differentiable $\mathbb{P}_n$-invariant convex function with $1 \in J$, and $F\colon J_0^n \to J^n$ be a $\mathbb{P}_n$-equivariant convex function such that $\nabla\Phi(\mathbf{e}) \geq 0$. Let $\mathbf{x}, \mathbf{y} \in J_0^n$.*

If
$$\mathbf{y} \prec \mathbf{x}, \tag{31}$$
then
$$\langle \nabla\Phi(\mathbf{e}), F(\mathbf{y}) - \mathbf{e}\rangle \leq \Phi(F(\mathbf{x})) - \Phi(\mathbf{e}). \tag{32}$$

Proof. Assumption (31) ensures that there exists an $n \times n$ doubly stochastic matrix $\mathbf{S}$ such that $\mathbf{y} = \mathbf{S}\mathbf{x}$. Simultaneously, $\mathbf{e} = \mathbf{S}^T\mathbf{e}$, as $\mathbf{S}^T$ is doubly stochastic, too.

It is now sufficient to make use of Theorem 2.3 for $\mathbf{a} = \mathbf{b} = \mathbf{e}$. □

Example 2.9. Let $\Phi\colon J^n \to \mathbb{R}$ be a differentiable $\mathbb{P}_n$-invariant convex function with $1 \in J$, and $F\colon J_0^n \to J^n$ be a $\mathbb{P}_n$-equivariant convex function such that $\nabla\Phi(\mathbf{e}) \geq 0$.

(i) Let $\mathbf{x} = (x_1, \ldots, x_n)^T \in J_0^n$ and $\mathbf{y} = \bar{\mathbf{x}} = (\bar{x}, \ldots, \bar{x})^T$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.
It is well known that $\bar{\mathbf{x}} \prec \mathbf{x}$. For this reason, Corollary 2.8 guarantees that
$$\langle \nabla\Phi(\mathbf{e}), F(\bar{\mathbf{x}}) - \mathbf{e}\rangle \leq \Phi(F(\mathbf{x})) - \Phi(\mathbf{e}).$$

(ii) Assume $J_0 = \mathbb{R}_+^n$.
Let $\mathbf{y} = (y_1, \ldots, y_n)^T \in \mathbb{R}_+^n$ and $\mathbf{x} = \mathbf{y}_s = (y_1 + \ldots + y_n, 0, \ldots, 0)^T \in \mathbb{R}_+^n$.
Since $\mathbf{y} \prec \mathbf{y}_s$, Corollary 2.8 gives
$$\langle \nabla\Phi(\mathbf{e}), F(\mathbf{y}) - \mathbf{e}\rangle \leq \Phi(F(\mathbf{y}_s)) - \Phi(\mathbf{e}).$$

Theorem 2.10. *Let $\varphi\colon J \to \mathbb{R}$ be a differentiable convex function, and $f\colon J_0 \to J$ be a convex function. Let $\mathbf{x} = (x_1, \ldots, x_n)^T \in J_0^n$, $\mathbf{y} = (y_1, \ldots, y_n)^T \in J_0^n$ and $\mathbf{a} = (a_1, \ldots, a_n)^T \in J^n$, $\mathbf{b} = (b_1, \ldots, b_n)^T \in J^n$ with $\varphi'(b_i) \geq 0$ for $i = 1, \ldots, n$.*

If
$$\mathbf{y} = \mathbf{S}\mathbf{x} \quad \text{and} \quad \mathbf{a} = \mathbf{S}^T\mathbf{b} \tag{33}$$
for some $n \times n$ doubly stochastic matrix $\mathbf{S}$, then
$$\sum_{i=1}^n \varphi'(b_i)(f(y_i) - b_i) \leq \sum_{i=1}^n (\varphi(f(x_i)) - \varphi(a_i)). \tag{34}$$

Proof. We introduce functions $\Phi\colon J^n \to \mathbb{R}$ and $F\colon J_0^n \to J^n$ of n-variables in the following manner:
$$\Phi(\mathbf{c}) = \sum_{i=1}^n \varphi(c_i) \quad \text{for } \mathbf{c} = (c_1, \ldots, c_n)^T \in J^n\ , \tag{35}$$
and
$$F(\mathbf{z}) = (f(z_1), \ldots, f(z_n))^T \quad \text{for } \mathbf{z} = (z_1, \ldots, z_n)^T \in J_0^n\ . \tag{36}$$

Then $\Phi\colon J^n \to \mathbb{R}$ is a differentiable $\mathbb{P}_n$-invariant convex function with
$$\nabla\Phi(\mathbf{c}) = (\varphi'(c_1), \ldots, \varphi'(c_n))^T \quad \text{for } \mathbf{c} = (c_1, \ldots, c_n)^T \in J^n\ .$$

Additionally, $F\colon J_0^n \to J^n$ is a $\mathbb{P}_n$-equivariant function convex with respect to the componentwise order $\leq$ on $\mathbb{R}^n$.

So, we are allowed to use Theorem 2.3. Therefore we deduce that

$$\sum_{i=1}^{n} \varphi'(b_i)(f(y_i) - b_i) = \langle \nabla\Phi(\mathbf{b}), F(\mathbf{y}) - \mathbf{b}\rangle \leq \Phi(F(\mathbf{x})) - \Phi(\mathbf{a})$$
$$= \sum_{i=1}^{n} (\varphi(f(x_i)) - \varphi(a_i)),$$

which proves (34). □

We now discuss conditions under which inequality (34) is a refinement of the standard gradient inequality.

Theorem 2.11. *Let $\varphi\colon J \to \mathbb{R}$ be a differentiable convex function on $J = (\alpha, \beta)$, such that φ is decreasing on $J_1 = (\alpha, \gamma)$ and φ is nondecreasing on $J_2 = [\gamma, \beta)$. Assume $1 \in J$ with $\varphi'(1) \geq 0$.*

Let ψ be the restriction to J_1 of φ, and $\psi^{-1}\colon \psi(J_1) \to J_1$ the inverse function of ψ.

Let $g\colon J_0 \to \psi(J_1)$ be a concave function such that the composite function $f = \psi^{-1} \circ g\colon J_0 \to J_1$ is well-defined.

Let $\mathbf{x} = (x_1, \ldots, x_n)^T \in J_0^n$, $\mathbf{y} = (y_1, \ldots, y_n)^T \in J_0^n$ be such that

$$\mathbf{y} \prec \mathbf{x}. \tag{37}$$

Then

$$\sum_{i=1}^{n} \varphi'(1)(f(y_i) - 1) \leq \sum_{i=1}^{n} (\varphi(f(x_i)) - \varphi(1)) \leq \sum_{i=1}^{n} (\varphi(f(y_i)) - \varphi(1)). \tag{38}$$

Proof. The function $f\colon J_0 \to J_1 \subset J$ is convex on J_0. In fact, g is concave and ψ^{-1} is nonincreasing and convex.

Therefore the left-hand side inequality of (38) follows easily from (34).

We shall focus on the proof of the right-hand side inequality of (38).

It is not hard to check that $\varphi \circ f = \psi \circ f = \psi \circ \psi^{-1} \circ g = g$.

So, the composition $\varphi \circ f\colon J_0 \to \mathbb{R}$ is concave on J_0, because g is so.

We introduce functions $\Phi\colon J^n \to \mathbb{R}$ and $F\colon J_0^n \to J^n$ of n-variables as in (35)–(36).

In consequence, for $\mathbf{z} = (z_1, \ldots, z_n)^T \in J_0^n$,

$$\Phi(F(\mathbf{z})) = \Phi((f(z_1), \ldots, f(z_n))^T) = \sum_{i=1}^{n} \varphi(f(z_i)) = \sum_{i=1}^{n} g(z_i).$$

Since g is concave on J_0, by virtue of Theorem **A** we infer that the function $\Phi \circ F$ is Schur-concave on J_0^n. For this reason, on account of (37), we get

$$\Phi(F(\mathbf{x})) \leq \Phi(F(\mathbf{y})). \tag{39}$$

Again, by (37), we obtain $\mathbf{y} = \mathbf{S}\mathbf{x}$ for some $n \times n$ doubly stochastic matrix $\mathbf{S}$.

We now define $\mathbf{b} = \mathbf{e}$ and $\mathbf{a} = \mathbf{S}^T\mathbf{b}$, where $\mathbf{e} = (1, \dots, 1)^T \in \mathbb{R}^n$. So, we get $\mathbf{a} = \mathbf{S}^T\mathbf{e} = \mathbf{e}$. Thus we have $\Phi(\mathbf{a}) = \Phi(\mathbf{b})$.

Finally, we conclude from (39) that

$$\Phi(F(\mathbf{x})) - \Phi(\mathbf{e}) \le \Phi(F(\mathbf{y})) - \Phi(\mathbf{e}).$$

This is equivalent to the right-hand side inequality of (38), which was to be proved.

□

Example 2.12. We consider the function $\varphi(t) = t \log t$ for $t \in J = (0, \infty)$. It is differentiable with $\varphi'(t) = \log t + 1$ for $t \in J$. Hence, φ' is a nondecreasing function on J, so φ is convex on J.

In particular, $\varphi'(t) < 0$ for $t \in J_1 = (0, e^{-1})$, and $\varphi'(t) \ge 0$ for $t \in J_2 = [e^{-1}, \infty)$. So, φ is decreasing on J_1, and φ is nondecreasing on J_2.

Let ψ be the restriction to J_1 of φ, and $\psi^{-1}\colon \psi(J_1) \to J_1$ be the inverse function of ψ. Since ψ is decreasing on J_1 and because we have $\lim_{t \to 0^+} \psi(t) = 0$, we get $\psi(J_1) = (-e^{-1}, 0)$.

Let $g\colon J_0 \to \psi(J_1) = (-e^{-1}, 0)$ be a concave function. Then the composite function $f = \psi^{-1} \circ g\colon J_0 \to J_1 = (0, e^{-1})$ is well-defined.

Fix arbitrarily $\mathbf{x} = (x_1, \dots, x_n)^T \in J_0^n$ and $\mathbf{y} = (y_1, \dots, y_n)^T \in J_0^n$ such that

$$\mathbf{y} \prec \mathbf{x}. \tag{40}$$

Set $a_i = b_i = 1$ for $i = 1, \dots, n$. Hence $\varphi(b_i) = \varphi(1) = 0$ and $\varphi'(b_i) = \varphi'(1) = 1$.

In this situation (38) takes the form

$$\sum_{i=1}^{n} (f(y_i) - 1) \le \sum_{i=1}^{n} f(x_i) \log f(x_i) \le \sum_{i=1}^{n} f(y_i) \log f(y_i). \tag{41}$$

The outer inequality of (41) comes from the standard gradient inequality for Φ induced by $\varphi(t) = t \log t$, $t > 0$, at the points $F(\mathbf{y})$ and $\mathbf{e}$, where Φ and F are defined as in (35)–(36).

References

[1] M. Adil Khan, S. Ivelić Bradanović, J. Pečarić: *On Sherman's type inequalities for n-convex function with applications*, Konuralp J. Math. 4 (2016) 255–260.

[2] M. Adil Khan, J. Khan, J. Pečarić: *Generalizations of Sherman's inequality by Montgomery identity and Green function*, Electron. J. Math. Analysis Appl. 5(1) (2017) 1–16.

[3] R. P. Agarwal, S. Ivelić Bradanović, J. Pečarić: *Generalizations of Sherman's inequality by Lidstone's interpolating polynomial*, J. Inequal. Appl. 2016(6) (2016) 18 pp.

[4] A.-M. Burtea, *Two examples of weighted majorization*, An. Univ. Craiova, Ser. Mat. Inf. 37 (2010) 92–99.

[5] G. M. Hardy, J. E. Littlewood, G. Pólya: *Inequalities*, 2nd ed., Cambridge University Press, Cambridge (1952).

[6] S. Ivelić Bradanović, N. Latif, J. Pečarić: *Generalizations of Sherman's theorem by Taylor's formula*, J. Inequal. Special Functions 8 (2017) 18–30.

[7] S. Ivelić Bradanović, J. Pečarić: *Generalizations of Sherman's inequality*, Period. Math. Hung. 74 (2017) 197–219.

[8] J. Karamata: *Sur une inégalité relative aux fonctions convexes*, Publ. Math. Univ. Belgrade 1 (1932) 145–148.

[9] A. W. Marshall, I. Olkin, B. C. Arnold: *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., Springer, New York (2011).

[10] M. Niezgoda: *Remarks on Sherman like inequalities for* (α, β)*-convex functions*, Math. Inequal. Appl. 17 (2014) 1579–1590.

[11] M. Niezgoda: *On Sherman method to deriving inequalities for some classes of functions related to convexity*, in: *Advances in Mathematical Inequalities and Applications*, P. Agarwal, S. S. Dragomir, M. Jleli, B. Samet (eds.), Trends in Mathematics, Springer, Berlin (2018) 219–245.

[12] M. Niezgoda: *Nonlinear Sherman-type inequalities*, Adv. Nonlinear Anal. 9(1) (2020) 168–175.

[13] R. Rado: *An inequality*, J. London Math. Soc. 27 (1952) 1–6.

[14] S. Sherman: *On a theorem of Hardy, Littlewood, Pólya, and Blackwell*, Proc. Nat. Acad. Sci. USA 37 (1951) 826–831.