lions@epfl

EE-556: Mathematics of Data: From Theory to Computation
Laboratory for Information and Inference Systems
Fall 2018

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Instructor
Prof. Volkan Cevher

Head TA
Ya-Ping Hsieh

Recitation 3 - Solution Set

1 Convexity of functions and sets

Problem 1: Convexity definitions

(a) If f is convex, then by the given definition of convexity, we have $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$$
$$f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$

Adding the above two inequalities together yield the desired inequality.

For the opposite direction, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \forall t \in [0,1]$, f satisfies $t\langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0$. Hence,

$$\begin{aligned}
f(\mathbf{y}) &= f(\mathbf{x}) + \int_0^1 \frac{d}{dt} f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))dt \\
&= f(\mathbf{x}) + \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt \\
&= f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt \\
&\geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle
\end{aligned}$$

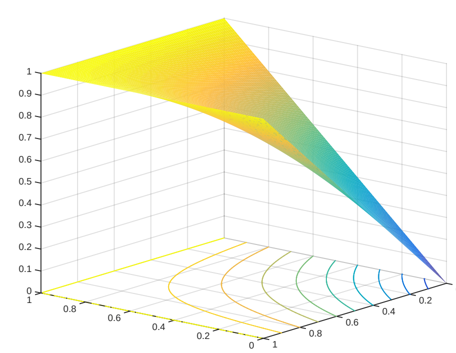
which implies that f is convex.

(b) For all $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \in \mathrm{epi}(f)$ and $\forall (\mathbf{x}, y) = \lambda(\mathbf{x}_1, y_1) + (1-\lambda)(\mathbf{x}_2, y_2)$ for some $\lambda \in [0,1]$, we have:

$$\begin{aligned}
f \text{ is convex} &\Leftrightarrow f(\mathbf{x}) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \\
&\Leftrightarrow f(\mathbf{x}) \leq \lambda y_1 + (1-\lambda) y_2 \\
&\Leftrightarrow (\mathbf{x}, y) \in \mathrm{epi}(f) \\
&\Leftrightarrow \mathrm{epi}(f) \text{ is convex}
\end{aligned}$$

Problem 2: Non-convex/concave function

```
x1 = 0:0.01:1;
x2 = 0:0.01:1;
[X, Y] = meshgrid(x1,x2);
f = @(x) 1 - prod(1-x);
for i=1:size(X,1)
    for j=1:size(X,2)
        Z(i,j) = f([X(i,j), Y(i,j)]);
    end
end
figure, surfc(X,Y,Z)
```



(a) $[\nabla f(\mathbf{x})]_1 = (1 - x_2), [\nabla f(\mathbf{x})]_2 = (1 - x_1)$. Then the stationary point is $(1,1)$ which is neither a maximum nor a minimum of f (actually the min/max of f are both unbounded). Hence, f cannot be convex/concave.

(b) We compute the hessian of f, $[\nabla^2 f(\mathbf{z})]_{ij} = -1$ if $i \neq j$, zero otherwise. Note that for $\mathbf{x} = \mathbf{y} + \mathbf{d}$ where $\mathbf{d} \in \mathbb{R}^{\mathbf{n}}_{+}$, we have $\frac{1}{2}(\mathbf{x} - \mathbf{y})^T \nabla^2 f(\mathbf{z})(\mathbf{x} - \mathbf{y}) = -d_1 d_2 \leq 0$. Hence by Taylor's theorem $f(\mathbf{x}) \leq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$, which implies that f is concave.

Problem 3: Binary logistic regression: Geometric properties of the objective function

(a) Define $f_i(\mathbf{x}) := \log\left[1 + \exp\left(-b_i \langle \mathbf{a}_i, x \rangle\right)\right]$. Using the chain rule in calculus, we obtain, for all $1 \leq k \leq p$,

$$\frac{\partial f_i}{\partial x_k}(\mathbf{x}) = \frac{-b_i \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right) (\mathbf{a}_i)_k}{1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}, \tag{1}$$

where $(\mathbf{a}_i)_k$ denotes the k-th element of $\mathbf{a}_i$. Therefore,

$$\frac{\partial f}{\partial x_k}(\mathbf{x}) = \frac{\partial \sum_{i=1}^n f_i}{\partial x_k}(\mathbf{x}) = \sum_{i=1}^n \frac{\partial f_i}{\partial x_k}(\mathbf{x}) = -\sum_{i=1}^n \frac{-b_i \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right) (\mathbf{a}_i)_k}{1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}.$$

By the fact that

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \frac{\partial f}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f}{\partial x_p}(\mathbf{x}) \end{bmatrix}^{\mathrm{T}},$$

we have

$$\nabla f(\mathbf{x}) = -\sum_{i=1}^n \frac{-b_i \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}{1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)} \mathbf{a}_i.$$

(b) By (1) and the chain rule in calculus, we obtain, for all $1 \leq k, \ell \leq p$,

$$\begin{aligned}
\frac{\partial^2 f_i}{\partial x_k \partial x_\ell}(\mathbf{x}) &= \frac{\partial}{\partial x_k}\left[\frac{-b_i \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right) (\mathbf{a}_i)_\ell}{1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}\right] \\
&= \frac{\partial}{\partial x_k}\left[\frac{-b_i (\mathbf{a}_i)_\ell}{1 + \exp\left(b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}\right] \\
&= \frac{(b_i)^2 \exp\left(b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}{\left[1 + \exp\left(b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)\right]^2} (\mathbf{a}_i)_k (\mathbf{a}_i)_\ell \\
&= \frac{\exp\left(b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}{\left[1 + \exp\left(b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)\right]^2} (\mathbf{a}_i)_k (\mathbf{a}_i)_\ell \\
&= \frac{\exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}{\left[1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)\right]^2} (\mathbf{a}_i)_k (\mathbf{a}_i)_\ell.
\end{aligned}$$

Then, as in the derivation of ∇f,

$$\frac{\partial^2 f}{\partial x_k \partial x_\ell}(\mathbf{x}) = \sum_{i=1}^n \frac{\exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}{\left[1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)\right]^2} (\mathbf{a}_i)_k (\mathbf{a}_i)_\ell.$$

Recall that the (k, ℓ)-th element of $\nabla^2 f(\mathbf{x})$ is given by

$$\left[\nabla^2 f(\mathbf{x})\right]_{k,\ell} = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(\mathbf{x}).$$

We obtain

$$\nabla^2 f(\mathbf{x}) = \sum_{i=1}^n \frac{\exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)}{\left[1 + \exp\left(-b_i \langle \mathbf{a}_i, \mathbf{x} \rangle\right)\right]^2} \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}}.$$

(c) Recall that a matrix $\mathbf{M} \in \mathbb{R}^{p \times p}$ is non-invertible if it has a zero eigenvalue. Moreover, $\mathbf{M}$ has a zero eigenvalue if there exists some $\mathbf{v} \in \mathbb{R}^p$ (eigenvector), such that $\mathbf{M}\mathbf{v} = 0\mathbf{v} = \mathbf{0}$. Regarding the expression of $\nabla^2 f(\mathbf{x})$ we just obtained, it suffices to consider a matrix $\mathbf{M}$ of the form

$$\mathbf{M} = \sum_{i=1}^n \gamma_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}},$$

for some positive numbers $\gamma_1, \gamma_2, \ldots, \gamma_n$. Notice that $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ spans an n-dimensional subspace $E \subset \mathbb{R}^p$. When $n < p$, we can find a non-empty $(p - n)$-dimensional subspace $E' \subset \mathbb{R}^p$ orthogonal to E. Choosing $\mathbf{v} \in E'$, we have

$$\mathbf{M}\mathbf{v} = \sum_{i=1}^n \gamma_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}} \mathbf{v} = \mathbf{0}.$$

Hence if $n < p$, $\mathbf{M}$ is non-invertible.

(d) Recall that f is convex, if $\nabla^2 f(\mathbf{x})$ is positive semi-definite for all $\mathbf{x} \in \mathbb{R}^p$. Also recall that $\nabla^2 f(\mathbf{x})$ is positive semi-definite, if $\mathbf{v}^{\mathrm{T}}\nabla^2 f(\mathbf{x})\mathbf{v} \geq 0$ for any $\mathbf{v} \in \mathbb{R}^p$. As in the previous question, it suffices to consider $\mathbf{M}$; then $\nabla^2 f(\mathbf{x})$ becomes a special case. We write

$$\mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v} = \mathbf{v}^{\mathrm{T}}\left(\sum_{i=1}^{n} \gamma_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}}\right)\mathbf{v} = \sum_{i=1}^{n} \gamma_i \mathbf{v}^{\mathrm{T}}\mathbf{a}_i\mathbf{a}_i^{\mathrm{T}}\mathbf{v} = \sum_{i=1}^{n} \gamma_i \left(\mathbf{a}_i^{\mathrm{T}}\mathbf{v}\right)^2 \geq 0.$$

Hence, $\nabla^2 f(\mathbf{x})$ is positive semi-definite, implying that f is convex.

Problem 4: Functions of Complex Variables

(a) We write

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{y}, \mathbf{y} \rangle + \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x} \rangle - 2\langle \mathbf{y}, \mathbf{A}\mathbf{x} \rangle \\ &= \langle \mathbf{y}, \mathbf{y} \rangle + \left\langle A^{\mathrm{T}}\mathbf{A}\mathbf{x}, \mathbf{x} \right\rangle - 2\left\langle \mathbf{A}^{\mathrm{T}}\mathbf{y}, \mathbf{x} \right\rangle \\ &= \langle \mathbf{y}, \mathbf{y} \rangle + \sum_{i,j} (\mathbf{A}^{\mathrm{T}}\mathbf{A})_{i,j} x_i x_j - 2\sum_i (\mathbf{A}^{\mathrm{T}}\mathbf{y})_i x_i. \end{aligned}$$

Then for any $1 \leq i \leq p$,

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = 0 + 2\sum_j (\mathbf{A}^{\mathrm{T}}\mathbf{A})_{i,j} x_j - 2(\mathbf{A}^{\mathrm{T}}\mathbf{y})_i = 2(\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x})_i - (\mathbf{A}^{\mathrm{T}} y)_i.$$

Therefore,

$$\nabla f(\boldsymbol{x}) := \left[\frac{\partial f(\boldsymbol{x})}{\partial(x_1)}, \frac{\partial f(\boldsymbol{x})}{\partial(x_2)}, \dots, \frac{\partial f(\boldsymbol{x})}{\partial(x_p)}\right]^{\mathrm{T}} = -2\mathbf{A}^{\mathrm{T}}(\mathbf{y} - \mathbf{A}\mathbf{x}).$$

(b) If φ is a non-constant real-valued function, then we have $v(x, y) \equiv 0$, and hence

$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} \equiv 0.$$

However, $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ are not always zero, as φ is non-constant, so the Cauchy-Riemann equations are violated. Therefore, one cannot write $\frac{\partial \varphi}{\partial z}$, as this is undefined. For any given $x_2, x_3, \dots, x_p \in \mathbb{C}$, define

$$\psi_1(x) := f([x, x_2, x_3, \dots, x_p]^{\mathrm{T}}),$$

for any $x \in \mathbb{C}$. Then ψ_1 is a non-constant real-valued function of a complex variable, and hence, by our argument above, is not differentiable; that is, one cannot write $\frac{\partial \psi_1}{\partial x} = \frac{\partial f(\mathbf{x})}{\partial x_1}$. Similarly, one cannot write $\frac{\partial f(\mathbf{x})}{\partial x_i}$ for any $1 \leq i \leq p$, meaning that the standard definition of ∇f is not applicable.

(c) We first verify the convexity of f (instead of $\tilde{f}$). For any $\mathbf{u}, \mathbf{v} \in \mathbb{C}^p$ and $\alpha \in (0, 1)$, we write

$$\begin{aligned} & f((1-\alpha)\mathbf{u} + \alpha\mathbf{v}) \\ &\quad = \|\mathbf{y} - \mathbf{A}[(1-\alpha)\mathbf{u} + \alpha\mathbf{v}]\|_2^2 \\ &\quad = \|[(1-\alpha)\mathbf{y} + \alpha\mathbf{y}] - \mathbf{A}[(1-\alpha)\mathbf{u} + \alpha\mathbf{v}]\|_2^2 \\ &\quad = (1-\alpha)^2\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \alpha^2\|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 + 2\alpha(1-\alpha)\langle \mathbf{y} - \mathbf{A}\mathbf{u}, \mathbf{y} - \mathbf{A}\mathbf{v} \rangle \\ &\quad = (1-\alpha)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \alpha\|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 - \alpha(1-\alpha)\|(\mathbf{y} - \mathbf{A}\mathbf{u}) - (\mathbf{y} - \mathbf{A}\mathbf{v})\|_2^2 \\ &\quad \leq (1-\alpha)\|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 + \alpha\|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 \\ &\quad = (1-\alpha)f(\mathbf{u}) + \alpha f(\mathbf{v}); \end{aligned} \tag{2}$$

hence f is convex. Setting $\mathbf{x} = [\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}, \mathbf{u}_{\mathrm{I}}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{y} = [\mathbf{v}_{\mathrm{R}}^{\mathrm{T}}, \mathbf{v}_{\mathrm{I}}^{\mathrm{T}}]^{\mathrm{T}}$, we obtain by (2) that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{2p}$ and $\alpha \in (0, 1)$,

$$\tilde{f}((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1-\alpha)\tilde{f}(\mathbf{x}) + \alpha\tilde{f}(\mathbf{y}),$$

which verifies the convexity of $\tilde{f}$.

Since $\mathcal{X}$ is convex, we have, for any $\mathbf{u}, \mathbf{v} \in \mathcal{X}$ and $\alpha \in (0, 1)$,

$$(1-\alpha)\mathbf{u} + \alpha\mathbf{v} \in \mathcal{X}. \tag{3}$$

Setting $\mathbf{x} = [\mathbf{u}_{\mathrm{R}}^{\mathrm{T}}, \mathbf{u}_{\mathrm{I}}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{y} = [\mathbf{v}_{\mathrm{R}}^{\mathrm{T}}, \mathbf{v}_{\mathrm{I}}^{\mathrm{T}}]^{\mathrm{T}}$, we obtain by (3) that

$$(1-\alpha)\mathbf{x} + \alpha\mathbf{y} \in \tilde{\mathcal{X}},$$

which verifies the convexity of $\tilde{\mathcal{X}}$.

(d) We write

$$\tilde{f}(\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}) = \left\|(\boldsymbol{y}_{\mathrm{R}} + \mathrm{i}\boldsymbol{y}_{\mathrm{I}}) - (\boldsymbol{A}_{\mathrm{R}} + \mathrm{i}\boldsymbol{A}_{\mathrm{I}})(\boldsymbol{x}_{\mathrm{R}} + \mathrm{i}\boldsymbol{x}_{\mathrm{I}})\right\|_2^2$$
$$= \underbrace{\left\|\boldsymbol{y}_{\mathrm{R}} - \boldsymbol{A}_{\mathrm{R}}\boldsymbol{x}_{\mathrm{R}} + \boldsymbol{A}_{\mathrm{I}}\boldsymbol{x}_{\mathrm{I}}\right\|_2^2}_{\tilde{f}_1(\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix})} + \underbrace{\left\|\boldsymbol{y}_{\mathrm{I}} - \boldsymbol{A}_{\mathrm{R}}\boldsymbol{x}_{\mathrm{I}} - \boldsymbol{A}_{\mathrm{I}}\boldsymbol{x}_{\mathrm{R}}\right\|_2^2}_{\tilde{f}_2(\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix})}.$$

Let us try to reuse the results for the first question, by writing

$$\tilde{f}_1(\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}) = \left\|\boldsymbol{y}_{\mathrm{R}} - \begin{bmatrix} \boldsymbol{A}_{\mathrm{R}} & -\boldsymbol{A}_{\mathrm{I}} \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}\right\|_2^2.$$

Then according to the answer to the first question, we have

$$\nabla\tilde{f}_1(\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}) = -2\begin{bmatrix} \boldsymbol{A}_{\mathrm{R}}^{\mathrm{T}} \\ -\boldsymbol{A}_{\mathrm{I}}^{\mathrm{T}} \end{bmatrix}\left(\boldsymbol{y}_{\mathrm{R}} - \begin{bmatrix} \boldsymbol{A}_{\mathrm{R}} & -\boldsymbol{A}_{\mathrm{I}} \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}\right).$$

Similarly, we obtain

$$\nabla\tilde{f}_2(\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}) = -2\begin{bmatrix} \boldsymbol{A}_{\mathrm{I}}^{\mathrm{T}} \\ \boldsymbol{A}_{\mathrm{R}}^{\mathrm{T}} \end{bmatrix}\left(\boldsymbol{y}_{\mathrm{I}} - \begin{bmatrix} \boldsymbol{A}_{\mathrm{I}} & \boldsymbol{A}_{\mathrm{R}} \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix}\right).$$