Problem 9.1.

a) Let $v, w \in \mathcal{M}$. Let $t \in [0,1]$. f is convex hence

$$f(tv + (1-t)w) \leq \underbrace{t f(v) + (1-t) f(w)}_{= t m + (1-t) m = m}$$

because v, w are minimizers of f.
By definition of m, we have $m \leq f(tv+(1-t)w)$.

We conclude that $f(tv+(1-t)w) = m$:
$tv + (1-t)w \in \mathcal{M}$, hence $\mathcal{M}$ is convex.

b) By contradiction, assume that there exists $v, w \in \mathcal{M}$ such that $v \neq w$.
By strict convexity of f:
$$f\left(\frac{1}{2}v + \frac{1}{2}w\right) < \frac{1}{2} f(v) + \frac{1}{2} f(w) = m,$$

which contradicts the definition of m.

Problem 9.2

a) Let $x \in \mathbb{R}^m$. | $\nabla f(x) = 2Mx + b$
$H_f(x) = 2M$.

f is twice differentiable. Hence, f is convex if and only if for all $x \in \mathbb{R}^m$, $H_f(x) = 2M$ is positive semidefinite, that is, if and only M is positive semidefinite.

b) f is convex, differentiable. Therefore:

f admits a minimizer $\iff$ there exists $x \in \mathbb{R}^n$ such that $\nabla f(x) = 0$

$\iff$ ________________, $2Mx + b = 0$

$\iff$ there exists $u \in \mathbb{R}^n$, such that

$$b = Mu$$

$\iff$ $b \in \mathrm{Im}(M)$.

Problem 9.3.

a) We start by showing that $\left| \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R} \\ x \longmapsto x^2 \end{array} \right.$ is strictly convex. Let $x, y \in \mathbb{R}$ and $t \in (0,1)$, and assume that $x \neq y$. Compute:

$(tx + (1-t)y)^2 = t^2x^2 + 2t(1-t)xy + (1-t)^2y^2.$

Hence $(tx + (1-t)y)^2 - (tx^2 + (1-t)y^2)$

$= -t(1-t)x^2 + 2t(1-t)xy - t(1-t)y^2$

$= -t(1-t)(x-y)^2 < 0.$

Which proves that $x \mapsto x^2$ is strictly convex.

Problem 9.3.

a) We start by showing that $\begin{array}{l} \mathbb{R} \to \mathbb{R} \\ x \mapsto x^2 \end{array}$ is strictly convex.

Let $x, y \in \mathbb{R}$ and $t \in [0,1]$. Compute:

$$\begin{aligned}(tx+(1-t)y)^2 - tx^2 - (1-t)y^2 &= t^2x^2 + 2t(1-t)xy + (1-t)^2y^2 - tx^2 - (1-t)^2y^2 \\ &= -(1-t)tx^2 + 2t(1-t)xy - t(1-t)y^2 \\ &= -t(1-t)(x-y)^2\end{aligned}$$

which is ≤ 0 and < 0 if $t \in (0,1)$ and $x \neq y$.
Hence $x \mapsto x^2$ is strictly convex.

Let us show now that $h(x) = \|x\|^2$ is strictly convex
Let $x, y \in \mathbb{R}^n$ and $t \in [0,1]$.

$$\begin{aligned}h(tx+(1-t)y) &= \sum_{i=1}^{n} (tx_i + (1-t)y_i)^2 \\ &\leq t\sum_{i=1}^{n} x_i^2 + (1-t)\sum_{i=1}^{n} y_i^2 = th(x) + (1-t)h(y)\end{aligned}$$

using what we proved above, we know that moreover the inequality is strict ~~when~~ if $0 < t < 1$ and $x \neq y$.
Hence h is strictly convex.

Now, if f is strongly convex, there exists a convex function g and $\alpha > 0$ such that $f = g + \alpha h$.
Since $\alpha > 0$ and h strictly convex, we get that f is strictly convex.

b) For $\alpha > 0$ we write $f_\alpha(x) = \varphi(x) - \frac{\alpha}{2}\|x\|^2$.
Notice that f_α is twice differentiable, with Hessian given by:

$$H_{f_\alpha}(x) = H_\varphi(x) - \alpha\, Id_n.$$

φ is strongly convex.
$\Leftrightarrow$ there exist $\alpha > 0$ such that f_α convex.
$\Leftrightarrow$ there exist $\alpha > 0$ such that for all $x \in \mathbb{R}^n$ $H_{f_\alpha}(x)$ is positive semidefinite (ie, has non-negative eigenvalues)

$\Leftrightarrow$ there exist $\alpha > 0$ such that for all $x \in \mathbb{R}^n$ the eigenvalues of $H_\varphi(x)$ are all $\geq \alpha$.

The last equivalence comes from the fact that the eigenvalues of $H_\varphi(x) - \alpha Id_n$ are the eigenvalues of $H_{\varphi(x)}$ minus α.

Problem 9.4
a) $f(x) = \|Ax - y\|^2 = (Ax-y)^T(Ax-y)$
$= x^T A^T A x - 2y^T A x + \|y\|^2.$
From 9.2 we know that
$$\nabla f(x) = 2A^TAx - 2A^Ty$$
and $H_f(x) = 2A^TA$.

Since we know that A^TA is positive semidefinite (because $v^TA^TAv = \|Av\|^2 \geq 0$ for all $v \in \mathbb{R}^n$)

we get that f is convex.

b) By the rank-nullity theorem:

$\dim \operatorname{Ker} A = m - \operatorname{rank}(A) \geq 1$ if $\operatorname{rank}(A) < m$.

In that case, we can find a non-zero vector v in the kernel of A.

Let now $x \in \mathbb{R}^m$ and $t \in (0,1)$.

$$f(tx + (1-t)(x+v)) = f(x + (1-t)v)$$
$$= \| Ax + (1-t)\underbrace{Av}_{=0} - y \|^2$$
$$= f(x)$$

On the other hand, we have that $f(x+v) = f(x)$, because $v \in \operatorname{Ker}(A)$. Hence:

$$f(tx + (1-t)(x+v)) = t f(x) + (1-t) f(x+v).$$

$v \neq 0$, hence $x \neq x+v$. We conclude that f is not strictly convex if $\operatorname{rank}(A) < m$.

c) If $\operatorname{rank}(A) = m$, then $\operatorname{rank}(A^T A) = \operatorname{rank}(A) = m$.

$A^T A$ is $m \times m$, we get that $A^T A$ is invertible.

Let $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of $A^T A$.

$A^T A$ is positive semi definite: $\lambda_1 \geq 0$.

$A^T A$ is invertible, hence $\lambda_1 > 0$.

We conclude that f is strongly convex because for all $x \in \mathbb{R}^n$, the eigenvalues of $H_f(x) = 2 A^T A$ are all $\geqslant 2\lambda_1 > 0$.

Problem 9.5

Let $x \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$ be fixed.

Define, for $t \in \mathbb{R}$, $g(t) = f(x+th) - f(x) - t\langle \nabla f(x), h \rangle$.

We will show that $\frac{\gamma}{2}\|h\|^2 \leq g(1) \leq \frac{L}{2}\|h\|^2$.

First, notice that
- $g(0) = 0$
- $g'(0) = 0$.

Second, compute, for $t \in \mathbb{R}$, $g''(t) = h^T Hf(x+th)\, h$

Since $\lambda_{max}(Hf(x+th)) = \max\limits_{\|v\|=1} v^T Hf(x+th)\, v$

and $\lambda_{min}(Hf(x+th)) = \min\limits_{\|v\|=1} v^T Hf(x+th)\, v$, because

$Hf(x+th)$ is symmetric, we get:

$$\lambda_{min}(Hf(x+th))\|h\|^2 \leq g''(t) \leq \lambda_{max}(Hf(x+th))\|h\|^2$$

which gives $\gamma\|h\|^2 \leq g''(t) \leq L\|h\|^2$, for all $t \in \mathbb{R}$.

By integrating these inequalities over $[0,t]$, for some $t \geq 0$ we get:

$$\gamma\|h\|^2 t \leq \int_0^t g''(s)\,ds \leq L\|h\|^2 t$$

$$\int_0^t g''(s)\,ds = g'(t) - g'(0) = g'(t).$$

Hence, for all $t \geq 0$: $\gamma \|h\|^2 t \leqslant g'(t) \leqslant L\|h\|^2 t$.

Integrating between 0 and 1 finally gives:

$$\gamma \|h\|^2 \underbrace{\int_0^1 t^2 dt}_{= \frac{1}{2}} \leqslant \underbrace{\int_0^1 g'(t) dt}_{\substack{= g(1) - g(0) \\ = g(1)}} \leqslant L \|h\|^2 \underbrace{\int_0^1 t^2 dt}_{= \frac{1}{2}}$$

<u>Conclusion</u>: $\frac{1}{2} \gamma \|h\|^2 \leqslant g(1) \leqslant \frac{1}{2} L \|h\|^2$, which implies:

$$f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \gamma \|h\|^2 \leqslant f(x+h) \leqslant f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} L \|h\|^2$$