Justification of Assumptions for Preconditioned Gradient Descent

1 Equivalence of the Convexity Inequality (Eq. 11)

Claim 1. *The standard convexity inequality $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ is equivalent to its preconditioned form $f(y) \geq f(x) + \langle \nabla_P f(x), y - x \rangle_P$.*

Proof. To prove the equivalence, we only need to show that the inner product terms are equal:

$$\langle \nabla f(x), y - x \rangle = \langle \nabla_P f(x), y - x \rangle_P$$

We start from the right-hand side and use the definition of the P-inner product:

$$\langle \nabla_P f(x), y - x \rangle_P = (\nabla_P f(x))^T P (y - x)$$

Now, we transform the left-hand side by substituting the relationship $\nabla f(x) = P \nabla_P f(x)$:

$$\begin{aligned}
\langle \nabla f(x), y - x \rangle &= \langle P \nabla_P f(x), y - x \rangle \\
&= (P \nabla_P f(x))^T (y - x) && \text{(by definition of the standard inner product)} \\
&= (\nabla_P f(x))^T P^T (y - x) && \text{(property of transpose } (AB)^T = B^T A^T\text{)} \\
&= (\nabla_P f(x))^T P (y - x) && \text{(since } P \text{ is symmetric, } P^T = P\text{)}
\end{aligned}$$

The expressions are identical, which confirms that the two forms of the convexity inequality are equivalent. □

2 Derivation of L_P-Smoothness (Eq. 12)

Claim 2. *If f is L-smooth w.r.t. the Euclidean norm ($\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$), then its preconditioned gradient $\nabla_P f$ is L_P-smooth w.r.t. the P-norm, with $L_P = L \cdot \lambda_{\max}(P^{-1})$.*

Proof. Our goal is to bound $\|\nabla_P f(x) - \nabla_P f(y)\|_P$. We start by relating the P-norm to the Euclidean norm using the symmetric square root of P, denoted $P^{1/2}$.

$$\begin{aligned}
\|\nabla_P f(x) - \nabla_P f(y)\|_P &= \|P^{1/2}(\nabla_P f(x) - \nabla_P f(y))\| \\
&= \|P^{1/2} P^{-1}(\nabla f(x) - \nabla f(y))\| && \text{(substituting } \nabla_P f = P^{-1} \nabla f\text{)} \\
&= \|P^{-1/2}(\nabla f(x) - \nabla f(y))\|
\end{aligned}$$

Lemma 1 (Norm Equivalence)**.** *For any vector $v \in \mathbb{R}^n$, the following equality holds:*

$$\|v\|_P = \|P^{1/2} v\|$$

where $P^{1/2}$ is the unique symmetric positive definite square root of P.

Proof. The proof follows directly from the definitions. We start with the square of the P-norm:

$$\begin{aligned}
\|v\|_P^2 &= \langle v, v\rangle_P && \text{(Definition of norm from inner product)}\\
&= v^T P v && \text{(Definition of P-inner product)}\\
&= v^T (P^{1/2} P^{1/2}) v && \text{(Substitute } P = P^{1/2}P^{1/2}\text{)}\\
&= v^T (P^{1/2})^T P^{1/2} v && \text{(Since } P^{1/2} \text{ is symmetric)}\\
&= (P^{1/2} v)^T (P^{1/2} v) && \text{(Property of transpose)}\\
&= \|P^{1/2} v\|^2 && \text{(Definition of Euclidean norm)}
\end{aligned}$$

Taking the square root of both sides yields the desired result: $\|v\|_P = \|P^{1/2}v\|$. □

Using the property of the matrix operator norm ($\|Az\| \leq \|A\|_{op}\|z\|$):

$$\|P^{-1/2}(\nabla f(x) - \nabla f(y))\| \leq \|P^{-1/2}\|_{op} \cdot \|\nabla f(x) - \nabla f(y)\|$$

Now, we apply the initial L-smoothness assumption:

$$\leq \|P^{-1/2}\|_{op} \cdot L \cdot \|x - y\|$$

To complete the proof, we must relate $\|x - y\|$ to $\|x - y\|_P$:

$$\begin{aligned}
\|x - y\| = \|I(x - y)\| &= \|P^{-1/2}P^{1/2}(x - y)\| \\
&\leq \|P^{-1/2}\|_{op} \cdot \|P^{1/2}(x - y)\| && \text{(by operator norm property)}\\
&= \|P^{-1/2}\|_{op} \cdot \|x - y\|_P && \text{(by definition of P-norm)}
\end{aligned}$$

Combining these inequalities, we get:

$$\begin{aligned}
\|\nabla_P f(x) - \nabla_P f(y)\|_P &\leq \|P^{-1/2}\|_{op} \cdot L \cdot \left(\|P^{-1/2}\|_{op} \cdot \|x - y\|_P\right)\\
&= L \cdot (\|P^{-1/2}\|_{op})^2 \cdot \|x - y\|_P
\end{aligned}$$

The operator norm of an SPD matrix is its largest eigenvalue. Thus, $\|P^{-1/2}\|_{op}^2 = (\lambda_{\max}(P^{-1/2}))^2 = \lambda_{\max}(P^{-1})$. The smoothness constant L_P is therefore:

$$L_P = L \cdot \lambda_{\max}(P^{-1}) = \frac{L}{\lambda_{\min}(P)}$$

This proves that L-smoothness implies L_P-smoothness. □

3 Derivation of the P-Descent Lemma (Eq. 13)

Claim 3 (P-Descent Lemma)**.** *L_P-smoothness of the preconditioned gradient implies the following quadratic upper bound:*

$$f(y) \leq f(x) + \langle \nabla_P f(x), y - x\rangle_P + \frac{L_P}{2}\|y - x\|_P^2$$

Proof. We use the Fundamental Theorem of Calculus and the tools established above.

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle \, dt = \int_0^1 \langle \nabla_P f(x + t(y - x)), y - x \rangle_P \, dt$$

Subtracting $\langle \nabla_P f(x), y - x \rangle_P = \int_0^1 \langle \nabla_P f(x), y - x \rangle_P \, dt$ from both sides gives:

$$f(y) - f(x) - \langle \nabla_P f(x), y - x \rangle_P = \int_0^1 \langle \nabla_P f(x + t(y - x)) - \nabla_P f(x), y - x \rangle_P \, dt$$

Applying the Cauchy-Schwarz inequality for the P-inner product:

$$\leq \int_0^1 \| \nabla_P f(x + t(y - x)) - \nabla_P f(x) \|_P \cdot \| y - x \|_P \, dt$$

Using the L_P-smoothness assumption, $\| \nabla_P f(z) - \nabla_P f(x) \|_P \leq L_P \| z - x \|_P$:

$$\begin{aligned}
&\leq \int_0^1 (L_P \cdot \| (x + t(y - x)) - x \|_P) \cdot \| y - x \|_P \, dt \\
&= \int_0^1 (L_P \cdot t \cdot \| y - x \|_P) \cdot \| y - x \|_P \, dt \\
&= L_P \| y - x \|_P^2 \int_0^1 t \, dt = \frac{L_P}{2} \| y - x \|_P^2
\end{aligned}$$

Rearranging the terms yields the P-Descent Lemma:

$$f(y) \leq f(x) + \langle \nabla_P f(x), y - x \rangle_P + \frac{L_P}{2} \| y - x \|_P^2$$

This completes the derivation. □