SCOPT: Self-Concordant OPTimization

SCOPT is a MATLAB implementation of the proximal gradient, proximal quasi-Newton, proximal Newton, and path-following interior-point algorithms for solving composite convex minimization problems involving self-concordant and self-concordant-like functions:

\[
\min_{x \in \mathbb{R}^n} \left\{ F(x) := f(x) + g(x) \right\},
\]

where \( f \) is a self-concordant or self-concordant-like function, and \( g \) is a proper, closed and convex function. Briefly, the function \( f \) is called self-concordant (with the parameter \( M_f > 0 \)) if

\[
|\phi'''(t)| \leq M_f \phi''(t)^{3/2},
\]

where \( \phi(t) := f(x + tu) \) for \( x \in \text{dom}(f) \), \( u \in \mathbb{R}^m \), and \( x + tu \in \text{dom}(f) \). Examples: \( f(X) := -\log \det(X) \), \( f(x,t) := -\log(t^2 - \|x\|^2_2) \), and \( f(x) = -\sum_{i=1}^n \log(a_i^T x - b_i) \) are all self-concordant.

The function \( g \) is usually assumed that its proximal operator

\[
\text{prox}_g(x) := \arg \min_{u \in \mathbb{R}^n} \left\{ g(u) + (1/2)\|u - x\|^2 \right\}
\]

is “easy” to compute (e.g., in a closed form or in polynomial time). SCOPT consists of several algorithms customized for solving the following convex optimization problems (but not limited to those):

- **Sparse Inverse Covariance Estimation Problems**
  Example:

  \[
  \min_{X \succeq 0} \left\{ -\log \det(X) + \text{tr}(SX) + \rho \|\text{vec}(X)\|_1 \right\}.
  \]

- **Poisson Intensity Reconstruction Problems**
  Example:

  \[
  \min_{x \geq 0} \left\{ \sum_{i=1}^m a_i^T x - \sum_{i=1}^m b_i \log(a_i^T x + \mu_i) + \rho \text{TV}(x) \right\}.
  \]

- **Heteroscedastic LASSO problems**
  Example:

  \[
  \min_{x \geq 0} \left\{ \frac{1}{m} - \log(\sigma) + (1/(2n))\|X\beta - \sigma y\|^2 + \rho \|\beta\|_1 \right\}.
  \]

- **Sparse logistic and multinomial logistic regression problems**
Example:

$$\min_{\mathbf{x} \in \mathbb{R}^{m \times p}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[ \log \left( 1 + \sum_{j=1}^{m} \exp \left( (\mathbf{W}^{(j)})^T \mathbf{X}^{(i)} \right) \right) - \sum_{i=1}^{m} y_j^{(j)} (\mathbf{W}^{(j)})^T \mathbf{X}^{(i)} \right] + \rho \| \text{vec}(\mathbf{X}) \|_1 \right\},$$

where $y^{(j)}$ and $\mathbf{W}^{(j)}$ are input data, $j = 1, \ldots, N$. 