Convex optimization

- minimize $f(x)$ subject to $g_i(x) \leq 0$
- Linear inequalities:
- Positivity: $x_3 \geq 0 \Rightarrow g_3(x) = -x_3$
- If it were “maximize $f(x)$”:
- If it were “$g_i(x) \geq 0$”:
- $f, g_i$ are convex functions $g_i: \mathbb{R}^n \rightarrow \mathbb{R}^m$
Gaussian GMs and glasso

• Gaussian graphical model
  ‣ $X \sim N(0, Q^{-1})$ $P(X=x \mid Q) =$
  ‣ $Q \in S_+^n$

• MLE: $\max_Q \ln P(X_{1:m} \mid Q) =$
  ‣ $\max_Q m \ln |Q/2\pi| - \sum x_j^T Q x_j$ or
  ‣ $\max_Q m \ln |Q/2\pi| - \sum x_j^T Q x_j - \lambda \sum_{i \neq j} |Q_{ij}|$
Feasibility & optimality

- Feasible, infeasible, unbounded
- Optimal value, optimal point
  - supremum, infimum
  - may not achieve optimum
  - no local optima
- Optimality criterion:
  - \( \nabla f(x^*)^T(x-x^*) \geq 0 \quad \forall x \in C \)
Types of convex program

• In order of increasing generality:
  ‣ Linear program
  ‣ Quadratic program
  ‣ Second-order cone program
  ‣ Semidefinite program
  ‣ “other”
Transformations

- change of variables
- monotone xform of objective
- changing signs
- slack variables on linear inequalities
- optimize out some variables
- implicit constraints

Use to make a non-convex program convex, or to make a convex program easier to solve
Solving convex programs

- Ellipsoid, subgradient, interior point
- Separation oracles: QP, SOCP, SDP
Example: logistic regression

- Data $x_i \in \mathbb{R}, y_i \in \{-1, 1\}$

- $P(y_i = 1 \mid x_i, w, b) = s(w^T x_i + b)$

  $s(z) = 1/(1 + \exp(-z))$
Example: logistic regression

\[ \text{arg max}_{w, b} P(w, b) \ P(y \mid x, w, b) = \]

\[ \text{arg max}_{w, b} P(w, b) \ \prod_i P(y_i \mid x_i, w, b) = \]
Variant: logistic lasso
Ex: soap bubbles

• Q: Dip a bent paper clip into soapy water. What shape film will it make?
  ‣ A:
  • Write $h(x,y)$ for height, $(x,y) \in S$
  • Area is:
  • Discretize:
Soap bubbles

\[ \min_h \int_S \sqrt{1 + \|\nabla h\|^2} \, dx \, dy \]

\[ h(x, y) = h_0(x, y) \quad \forall (x, y) \in \partial S \]
Ex: manifold learning

• Given points $x_1, \ldots, x_m$

• Find points $z_1, \ldots, z_m$

• Preserving local geometry
  ‣ given neighbor edges $N$
  ‣ distances

• If we preserve distances, we also preserve angles:
Step 1: “embed” $\mathbb{R}^n$ into $\mathbb{R}^n$

- While preserving local distances, move points to make manifold as flat as possible
  - to flatten, have points “repel” one another

- max

s.t.
Step 2: reduce to $\mathbb{R}^d$

- Now that manifold is flat, just use PCA:
Maximizing variance

• max \quad s.t.

• max \quad s.t.
Summary

- Solve SDP to “embed” $\mathbb{R}^n$ into $\mathbb{R}^n$
- Use PCA to embed $\mathbb{R}^n$ into $\mathbb{R}^d$
- Called “semidefinite embedding” or “maximum variance unfolding”

- Problems?
Dual functions

- Arbitrary function $F(x)$
- Dual is: $F^*(y) =$

- For example: $F(x) = x^T x / 2$
- $F^*(y) =$
Why dual functions?

• Will need them for duality of convex programs
  ‣ which is then good for everything LP duality is good for: bounds on optimal value, aiding understanding of problem, …
  ‣ and also for primal-dual interior point algorithms

• Turn out to be related to other kinds of duality:
  ‣ dual norms
  ‣ geometric duality
Fenchel’s inequality

• Recall definition: $F^*(y) = \sup_x [x^T y - F(x)]$
Duality and subgradients

• Suppose $F(x) + F^*(y) - x^T y = 0$
Duality examples

• $\frac{1}{2} - \ln(-x)$

• $e^x$

• $x \ln(x) - x$
More examples

• $F(x) = x^TQx/2 + c^Tx$, $Q$ pos. def.: 

• $F(X) = -\ln |X|$, $X$ psd:
Indicator functions

• Recall: for a set $S$,

$$I_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

• E.g., $I_{[-1,1]}(x)$:
Duals of indicators

• $l_a(x)$, point a:

• $l_C(x)$, set C:
To take dual of a piecewise linear function, interchange slopes and cutpoints.
Another example

- $F(x) = 1_c(x)$
  - $C = [-1, 1]^2$
- $F^*(y) =$
Properties

- $F(x) \geq G(x)$
- $F^*(y) \geq G^*(y)$
- $F^*$ is closed, convex

- If $F$ is differentiable:
Working with dual functions

- $G(x) = F(x) + k$

- $G(x) = k \cdot F(x) \quad k > 0$

- $G(x) = F(x) + a^T x$
Working with dual functions

- $G(x_1, x_2) = F_1(x_1) + F_2(x_2)$
An odd-looking operation

• Definition: infimal convolution

• E.g., $F_1(x) = I_{[-1,1]}(x)$, $F_2(x) = |x|$
Infimal convolution example

- \( F_1(x) = I_{\leq 0}(x), \quad F_2(x) = x^2 \)
Dual of infimal convolution

- $G(x) = F_1(x) \, \square \, F_2(x)$
- $G^*(y) =$
- $G(x) = F_1(x) + F_2(x)$