A SUBGRADIENT DESCENT METHOD FOR MISSING DATA PROBLEMS

Russell Luke

1. University of Delaware and Universität Göttingen

November 6, 2009

(with Jonathan Borwein)
Given a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ full-rank with $0 < m < n$, solve

$$ (\mathcal{P}_0) \quad \begin{array}{ll}
\text{minimize} & \|x\|_0 \\
\text{subject to} & Ax = b
\end{array} $$

where $\|x\|_0 := \sum_j |\text{sign}(x_j)|$ with $\text{sign}(0) := 0$.

Combinatorial optimization problem (hard to solve).
Central Problem: $\ell_0$ minimization

Solve instead

$$(P_1) \quad \begin{array}{l}
\text{minimize} \quad \|x\|_1 \\
\text{subject to} \quad Ax = b
\end{array} \quad x \in \mathbb{R}^n$$

where $\|x\|_1$ is the usual $\ell_1$ norm.

Linear programming problem...easy to solve for small problems.

Try solving the problem when $x$ is a $512 \times 512$ image...
Application: Crystallography

Given

\[(\text{autocor. transfer function – ATF – w/ missing pixels})\]

Find

\[(\text{true ATF})\]
Application: Crystallography

For the **linear mapping** \( A : \mathbb{R}^n \to \mathbb{R}^m \ (m < n) \), solve

\[
\text{minimize} \quad \|x\|_1 \\
\text{subject to} \quad x \in C
\]

where \( C := \{ x \in \mathbb{R}^n \mid Ax = b \} \).

One could apply Douglas-Rachford

\[
x_{n+1} = \frac{1}{2} \left( R_{f_1} R_{f_2} + I \right) (x_n)
\]

where

\[
R_{f_j} x := 2 \operatorname{prox}_{\alpha, f_j} x - x
\]

for \( f_1(x) = \|x\|_1 \) and \( f_2(x) = \iota_C(x) \), and \( \alpha > 0 \) fixed.
Application: Crystallography

Great strategy for big problems, but it is (arbitrarily) slow to converge and accuracy is bad.
Motivation

Find a variational/geometrical interpretation of the Candes-Tao (2004) probabilistic criteria for when the solution to $\mathcal{P}_1$ is unique and exactly matches the true signal $x_*$ and, as a by-product, better methods for solving the underlying problem.
A SUBGRADIENT DESCENT METHOD FOR MISSING DATA PROBLEMS

Russell Luke

Dual Convex Regularization
Subgradient descent with exact linesearch
Computational Results
Conclusion
References

Outline

- Dual convex regularization
- A subgradient descent algorithm with exact linesearch
- Computational results
Outline

Dual Convex Regularization

Subgradient descent with exact linesearch

Computational Results

Conclusion

References
**Fenchel duality**

The **Fenchel conjugate** of $f$, denoted $f^* : X^* \rightarrow ]-\infty, +\infty]$ and defined by

$$f^*(x^*) = \sup_{x \in X} \{ \langle x^*, x \rangle - f(x) \}.$$ 

For the $\ell_1$ problem, the norm is proper, convex, lsc and $b \in \text{core} (\text{Adom } f)$, so strong Fenchel duality holds:

$$\inf_{x \in \mathbb{R}^n} \{ \| x \|_1 \mid Ax = b \} = \sup_{y \in \mathbb{R}^m} \{ \langle b, y \rangle - \| (A^* y) \|_1^* \}.$$ 

where

$$\| x^* \|_1^* := \begin{cases} 0 & x^* \in [-1, 1] \\ +\infty & \text{else} \end{cases}$$
Elementary observations

The dual to $(P_1)$ is

$$\begin{align*}
(D_1) \quad & \text{maximize} & b^T y \\
\quad & \text{subject to} & (A^* y)_j \in [-1, 1] \quad j = 1, 2, \ldots, n.
\end{align*}$$

The solution includes a vertex of the polyhedron described by the constraints. The uniqueness of solutions to the primal problem depend on whether or not solutions to the dual problem reside along the edges or faces of the dual polyhedron.

If a solution $\bar{x}$ to $(P_1)$ is unique, then

$$m \geq \{ \text{the number of active constraints in } (D_1) \} = \|\bar{x}\|_0.$$
Elementary observations

For the $\ell_0$ problem, the function is proper, lsc but not convex, so only weak Fenchel duality holds:

$$\inf_{x \in \mathbb{R}^n} \{ \|x\|_0 \mid Ax = b \} \geq \sup_{y \in \mathbb{R}^m} \{ \langle b, y \rangle - \| (A^* y) \|_0^* \}.$$ 

where

$$\| x^* \|_0^* := \begin{cases} 0 & x^* = 0 \\ +\infty & \text{else} \end{cases}$$
Elementary observations

In other words, the dual to $(\mathcal{P}_0)$ is

\[(\mathcal{D}_0) \quad \begin{align*}
\text{maximize} & \quad b^T y \\
\text{subject to} & \quad A^* y = 0.
\end{align*}
\]

The primal problem is a combinatorial optimization problem; the dual problem, however, is a linear program, which is finitely terminating.

In fact, the solution to the dual problem is trivial: $\bar{y} = 0$...

which tells us nothing useful about the primal problem.
Relax/Regularize the dual and either solve this directly, or solve the corresponding regularized primal problem, or a mixture of the two.
Regularization/Relaxation: Shifted Fermi-Dirac Entropy

For \( L, \epsilon > 0 \), define

\[
\begin{align*}
    f_{\epsilon,L}^*(x) &:= \\
    &\sum_{j=1}^{n} \left[ \epsilon \left( \frac{(L + x_j) \ln(L + x_j) + (L - x_j) \ln(L - x_j)}{2L \ln(2)} - \frac{\ln(L)}{\ln(2)} \right) \right] \\
    &\quad \text{for } x \in [-L, L]^n \\
    &:= +\infty \quad \text{for } \|x\|_\infty > L.
\end{align*}
\]

Then

\[
f_{\epsilon,L}^{**}(x) = \sum_{j=1}^{n} \left[ \frac{\epsilon}{\ln(2)} \ln \left( 4^{x_j L/\epsilon} + 1 \right) - x_j L - \epsilon \right].
\]

\( f^* \) is proper, convex and lsc, thus \( f^{***} = f^* \) and we define \( f := f^{**} \).
Regularization/Relaxation: Shifted Fermi-Dirac Entropy
Regularization/Relaxation: Shifted Fermi-Dirac Entropy

For $L > 0$ fixed, in the limit as $\epsilon \to 0$ we have

$$\lim_{\epsilon \to 0} f^*_{\epsilon,L}(y) = \begin{cases} 0 & y \in [-L, L] \\ +\infty & \text{else} \end{cases} \quad \text{and} \quad \lim_{\epsilon \to 0} f_{\epsilon,L}(x) = L|x|.$$ 

For $\epsilon > 0$ fixed we have

$$\lim_{L \to 0} f^*_{\epsilon,L}(x) = \begin{cases} 0 & y = 0 \\ +\infty & \text{else} \end{cases} \quad \text{and} \quad \lim_{L \to 0} f_{\epsilon,L}(x) := 0.$$
Regularization/Relaxation: Shifted Fermi-Dirac Entropy

- $\| \cdot \|_0$ and $f^*_{\epsilon_0}$ have the same conjugate;
- $\| \cdot \|_{0^*} \neq \| \cdot \|_0$ while $f^{**}_{\epsilon_0} = f^*_{\epsilon_0}$;
- $f_{\epsilon,L}$ and $f^*_{\epsilon,L}$ are convex and smooth on the interior of their domains for all $\epsilon, L > 0$. This is in contrast to metrics of the form $\left( \sum_j |x_j|^p \right)$ which are nonconvex for $p < 1$. 
Regularization/Relaxation: Shifted Fermi-Dirac Entropy

Solve

\[
(\mathcal{D}_{L,\epsilon}) \quad \min_{y \in \mathbb{R}^m} f^*_{L,\epsilon}(A^*y) - \langle b, y \rangle
\]

This is a convex optimization problem, so equivalently we solve

\[
0 \in A\partial f^*_{L,\epsilon}(A^*y) - b
\]
Outline

Dual Convex Regularization

**Subgradient descent with exact linesearch**

Computational Results

Conclusion

References
Solve

\[ 0 \in A\partial f^*_{L,0}(A^*y) - b \]

via subgradient descent:

\[ \text{given } y_-, \text{ choose } v_- \in \partial f^*_{L,0}(A^*y_-), \lambda_- > 0 \text{ and construct } y_+ \text{ as} \]

\[ y_+ = y_- + \lambda_- (b - Av_-). \]

Issues:

- how to choose \( v_- \in \partial f^*_{L,0}(A^*y_-) \)
- how to choose step length \( \lambda_- \).
\( \epsilon = 0: \text{Choose } \nu_- \in \partial f_{L,0}^*(A^*y_-) \)

Note that \( f_{L,0}^* = \nu_{[-L,L]^n} \) so

\[
\partial \nu_{[-L,L]^n}(x^*) = N_{[-L,L]}(x^*) \\
= \{ \nu \in \mathbb{R}^n \mid \nu_j \leq 0 (j \in J_-), \ nu_j \geq 0 (j \in J_+), \ nu_j = 0 (j \in J_0) \}
\]

where \( J_- = \{ j \in \mathbb{N} \mid x_j = -L \} \), \( J_+ = \{ j \in \mathbb{N} \mid x_j = L \} \) and \( J_0 = \{ j \in \mathbb{N} \mid x_j \in ]-L, L[ \} \).

Choose \( \nu_- \in N_{[-L,L]}(A^*y_-) \) to be the solution to

\[
\text{minimize } \nu \in \mathbb{R}^n \quad \frac{1}{2} \| b - Av \|^2 \\
\text{subject to } \nu \in N_{[-L,L]^n}(A^*y_-)
\]
\( \epsilon = 0: \text{Choose } v_- \in \partial f_{L,0}^*(A^*y_-) \)

\[
\begin{align*}
(P_{v_-}) \quad & \underset{v \in \mathbb{R}^n}{\text{minimize}} \quad & \frac{1}{2} \| b - Av \|^2 \\
& \text{subject to} \quad & v \in N_{[-L,L]^n}(A^*y_-)
\end{align*}
\]

Define \( B = \{ v \in \mathbb{R}^n \mid Av = b \} \). Reformulate:

\[
(P_{v_-}) \quad \underset{v \in \mathbb{R}^n}{\text{minimize}} \quad \frac{\beta}{2(1 - \beta)} \text{dist}^2(v, B) + \iota_{N_{[-L,L]^n}(A^*y_-)}(v).
\]
\[ \epsilon = 0: \text{Choose } \nu_- \in \partial f_{L,0}^*(A^*y_-) \]

Approximate averaged alternating proximal reflections: Choose \( \nu^{(0)} \in \mathbb{R}^n \). For \( \nu \in \mathbb{N} \) set

\[ \nu^{(\nu+1)} = \frac{1}{2} \left( R_1 \left( R_2 \nu^{(\nu)} + \epsilon_\nu \right) + \rho_\nu + \nu^{(\nu)} \right), \quad (3) \]

where

- \( R_1 x := 2 \text{prox}_{\frac{\beta}{2(1-\beta)} \text{dist}(v,B)^2} x - x \)
- \( R_2 x := 2 \text{prox}_{\nu L[-L,L]^n(A^*y_-)} x - x \)
- The sequences \( \{\epsilon_\nu\} \) and \( \{\rho_\nu\} \) are the errors at each iteration, and assumed to be summable.
A SUBGRADIENT DESCENT METHOD FOR MISSING DATA PROBLEMS

Russell Luke

Dual Convex Regularization

Subgradient descent with exact linesearch

Computational Results

Conclusion

References

\( \epsilon = 0: \) Choose \( v_- \in \partial f_{L,0}^{*}(A^*y_-) \)

(L. 2008) Algorithm 3 is equivalent to: **Inexact Relaxed Averaged Alternating Reflections** (L. 2005)
Choose \( v^{(0)} \in \mathbb{R}^n \) and \( \beta \in [1/2, 1] \). For \( \nu \in \mathbb{N} \) set

\[
\begin{align*}
v^{(\nu+1)} &= \frac{\beta}{2} \left( R_B \left( R_{N_{[-L,L]}^n}(A^*y_-) v^{(\nu)} + \epsilon_n \right) + \rho_n + v^{(\nu)} \right) \\
&\quad + (1 - \beta) \left( P_{N_{[-L,L]}^n}(A^*y_-) v^{(\nu)} + \frac{\epsilon_n}{2} \right). \\
\end{align*}
\]

(4)

where \( R_B := 2P_B - I \) and likewise for \( R_{N_{[-L,L]}^n}(A^*y_-) \).

The sequence \( \{v^{(\nu)}\}_{\nu=1}^\infty \) converges to \( \overline{v} \) where \( P_B \overline{v} \)
solves (\( \mathcal{P}_{v_-} \)) (L. 2008, Combettes 2004).
\[ \varepsilon = 0: \textbf{Choose } \lambda_- \]

\textbf{Exact line search:} choose the largest \( \lambda_- \) that solves

\[
\min_{\lambda \in \mathbb{R}_+} f^*_{\lambda,0}(A^* y_+ + A^* \lambda(b - Av_-))
\]

Note that \( f^*_{\lambda,0}(A^* y_+ + A^* \lambda(b - Av_-)) = 0 \) for all \( A^* y_+ + A^* \lambda(b - Av_-) \in [-L, L]^n \), so really we solve

\[
\min_{\lambda \in \mathbb{R}_+} -\lambda
\]

\((P_\lambda)\)

subject to

\[
\lambda(A^*(b - Av_-))_j \leq L - (A^* y_-)_j \\
\lambda(A^*(b - Av_-))_j \geq -L - (A^* y_-)_j \quad j = 1, \ldots, n
\]
\( \epsilon = 0: \text{Choose } \lambda_- \)

**Exact line search:** Define

\[
J_+ = \{ j \mid (A^*(b - Av_-))_j > TOL \}, \\
J_- = \{ j \mid (A^*(b - Av_-))_j < -TOL \}
\]

\[
\lambda_- = \min \left\{ \min_{j \in J_+} \left\{ \frac{(L - (A^*y_-)_j)}{(A^*(b - Av_-))_j} \right\}, \min_{j \in J_-} \left\{ \frac{(-L - (A^*y_-)_j)}{(A^*(b - Av_-))_j} \right\} \right\}
\]

Algorithm terminates when \( J_+ = J_- = \emptyset \).
Outline

Dual Convex Regularization

Subgradient descent with exact linesearch

Computational Results

Conclusion

References
Computational Results

The solution to the dual $\bar{y}$ mapped by $A^*$:

This is a **qualitative** solution to the primal problem: it tells us the location and sign of the nonzero elements of the signal.
Computational Results

The solution to the primal $\bar{x}$ determined by the solution to

$$(P_{\bar{y}}) \quad \min_{x \in \mathbb{R}^n} \quad \frac{1}{2} ||b - Ax||^2$$

subject to $x \in \mathcal{N}_{[-L,L]^n}(A^*\bar{y})$

where $\bar{y}$ is the solution to the dual problem. The $\ell_\infty$ error is $10^{-12}$. 
Computational Results

Observations:

- Inner iterations can be shown to be arbitrarily slow: the solution sets to the subproblems are not metrically regular and the indicator function \( \nu_{N[−L,L]^n} \) is not coercive in the sense of Lions.

- The **algorithm** fails when there are too few samples relative to the sparsity of the true solution.
Conclusion
Conclusion

Work in progress:

- Characterize the recoverability of the true solution in terms of the regularity of the subproblem

\[
\begin{align*}
(\mathcal{P}_{v_{\ast}}) & \quad \min_{v \in \mathbb{R}^n} \quad \frac{1}{2} \| b - Av \|^2 \\
\text{subject to} & \quad v \in N_{[-L,L]^n}(A^*y_{\ast})
\end{align*}
\]
Conclusion

Thanks
Outline

Dual Convex Regularization

Subgradient descent with exact linesearch

Computational Results

Conclusion

References
References


References II

Relaxed averaged alternating reflections for diffraction imaging.

Finding best approximation pairs relative to a convex and a prox-regular set in Hilbert space.

J. J. Moreau.
Proximité et dualité dans un espace Hilbertian.